PCT / CA 00/00533 #2

CA 00/00533

5 JUL 2000 (05.07.00)

RECD 21 JUL 2000

WIPO PCT

PA 248620

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

4

United States Patent and Trademark Office

May 15, 2000

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: ***60/132,878***
FILING DATE: ***May 05, 1999***

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

L. Edelen

L. EDELEN
Certifying Officer

Please type a plus sign (+) inside this box → [+]

Docket Number: 11168.93

PROVISIONAL APPLICATION FOR PATENT COVER SHEET (Small Entity)

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c).

INVENTOR(S)/APPLICANT(S)		
Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Jerry John Howard	ROEMER DAVISON BUSSEY	Montreal, Quebec, Canada Montreal, Quebec, Canada Westmount, Quebec, Canada

☐ *Additional inventors are being named on page 2 attached hereto*

TITLE OF THE INVENTION (280 characters max)

IDENTIFICATION OF THE CANDIDA ALBICANS ESSENTIAL FUNGAL SPECIFIC GENES CaKRE5, CaALR1 AND CaCDC24 AND USE THEREOF IN ANTIFUNGAL DRUG DISCOVERY

CORRESPONDENCE ADDRESS

Direct all correspondence to:

☐ Customer Number [ ] → *Place Customer Number Bar Code Label here*

OR

☒ Firm or Individual Name	Goudreau Gage Dubuc & Martineau Walker				
Address	Stock Exchange Tower, Suite 3400, P.O. Box 242				
Address	800 Place-Victoria				
City	Montreal	State	Quebec	ZIP	H4Z 1E9
Country	Canada	Telephone	514-397-7449	Fax	514-397-4382

ENCLOSED APPLICATION PARTS *(check all that apply)*

☒ Specification — *Number of Pages* 31 — ☒ Small Entity Statement

☒ Drawing(s) — *Number of Sheets* 9 — ☐ Other (specify) [ ]

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT *(check one)*

☒ A check or money order is enclosed to cover the filing fees — FILING FEE AMOUNT $75.00

☐ The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: [ ]

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE _____ Date 05/03/1999

TYPED or PRINTED NAME Gaétan Prince — REGISTRATION NO. *(if appropriate)* 33107

TELEPHONE 514-397-7609

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

SEND TO: Box Provisional Application, Assistant Commissioner for Patents, Washington, DC 20231

P19SMALL/REV04

TITLE OF THE INVENTION

IDENTIFICATION OF THE *CANDIDA ALBICANS* ESSENTIAL FUNGAL SPECIFIC GENES *CaKRE5*, *CaALR1* AND *CaCDC24* AND USE THEREOF IN ANTIFUNGAL DRUG DISCOVERY

FIELD OF THE INVENTION

The present invention relates to the identification of novel essential fungal specific genes isolated in the yeast pathogen, *Candida albicans*, specifically CaKRE5, CaALR1 and CaCDC24, and particularly to their structural and functional relatedness to their *Sacharomyces cerevisiae* counterparts. More specifically the invention relates to the use of *CaKRE5, CaALR1* and *CaCDC24* in fungal diagnosis and antifungal drug discovery.

BACKGROUND OF THE INVENTION

Opportunistic fungi, including *Candida albicans*, *Aspergillus fumigatus*, *Cryptococcus neoformans*, and *Pneumocystis carinii*, are a rapidly emerging class of microbial pathogens, which cause systemic fungal infection or "mycosis" in patients whose immune system is weakened. Candida spp. rank as the predominant genus of fungal pathogens, accounting for approx. 8% of all bloodstream infections in hospitals today. Alarmingly, the incidence of life-theatening *C. albicans* infections or "candidiasis" have risen sharply over the last two decades, and ironically, the single greatest contributing factor to the prevalence of mycosis in hospitals today is modern medicine itself. Standard medical practices such as organ transplantation, chemotherapy and radiation therapy, suppress the immune system and make patients highly susceptible to fungal infection. Modern diseases, most notoriously, AIDS, also contribute to

this growing occurrence of fungal infection. In fact, Pneumocystis carinii infection is the number one cause of mortality for AIDS victims.

Treatment of fungal infection is hampered by the lack of safe and effective antifungal drugs. Antimycotic compounds used today; namely polyenes (amphotericin B) and azole-based derivatives (fluconazole), are of limited efficacy due to the nonspecific toxicity of the former and emmerging resistance to the latter. Resistance to fluconazole has increased dramatically throughout the decade particularly in *Candida* and *Aspergillus* spp.

Clearly, new antimycotic compounds must be developed to combat fungal infection and resistance. Part of the solution depends on the ellucidation of new antifungal drug targets (ie. molecules who's chemical inactivation/disruption results in cell death) distinct from that of current antifungal drugs which act by inactivating membrane/ergosterol composition. The identification of genes expressing proteins essential to cell viability in a broad spectrum of fungi, and absent in humans, serve as novel antifungal drug targets to which rational drug screening can be employed. In this way, drug screening can identify specific antifungal compounds that inactivate essential and fungal-specific genes, thereby mimicking the validated effect of the gene disruption.

A major advance in the study of pathogenesis and antifungal drug development comes from genome sequencing projects recently completed for the bakers yeast *Saccharomyces cerevisiae* and recently under way in *C. albicans*. Although *S. cerevisiae* is not itself pathogenic, it is closely related taxonomically to opportunistic pathogens including *C. albicans*. Consequently, many of the genes identified and studied in *S. cerevisiae* lend valuable insight into the identification and functional analysis of homologous genes present in the wealth of

sequence information provided by the Stanford *C. albicans* genome project (http://candida.stanford.edu), accelerating the isolation of *C. albicans* genes which may participate in the process of pathogenicity and cell viability.

Another dramatic advance from which antifungal drug discovery will benefit comes from the *S. cerevisiae* gene disruption consortium, in which the entire genome is being systematically disrupted (http: // sequence-www.stanford edu/group/yeastdeletion project /) dentification of all essential genes in this organism will enable strong predictions to be made as to which genes in *C. albicans* are similarly essential for cell viability.

The Bussey laboratory is a prominent contributor to the *S. cerevisiae* functional genomics project and has begun to apply this information to identifying potential antifungal drug targets in *C. albicans* (1). We have continued this approach to clone additional genes known to be essential for viability in *S. cerevisiae* and directly test whether an identical phenotype is observed in *C. albicans*. Such genes which are found to be essential in *C. albicans* serve as validated antifungal drug targets and provide novel reagents in antifungal drug screening programs.

There thus remains a need to identify essential fungal specific genes in *Candida albicans* and to use such genes in the discovery of drugs specifically directed against fungal pathogens.

The present invention seeks to meet these and other needs.

The present description refers to a number of documents, the content of which is herein incorporated by reference.

SUMMARY OF THE INVENTION

The invention concerns essential fungal specific genes in *Candida albicans* and their use in antifungal drug discovery.

The present invention further relates to Identification and disruption of the *Candida albicans* fungal specific genes, *CaKRE5*, *CaALR1*, and *CaCDC24* revealing structural and functional relatedness to their *Saccharomyces cerevisiae* counterparts, and validates their utility in fungal diagnosis and antifungal drug discovery.

In acccordance with the present invention, full length clones of *CaKRE5, Ca CDC24* and *CaALR1* using available fragments of *C. albicans* DNA were isolated by Polymerase Chain Reaction (PCR) to amplify genomic DNA derived from *C. albicans*. The PCR products were radiolabeled and used to probe the *C. albicans* genomic library by colony hybridization. DNA sequencing revealed complete open reading frames of CaKRE5, Ca CDC24 and CaALR1 sharing statistically significant homology to their *S. Cerevisiae* counterparts namely KRE5, CDC24 and ALR1 all of which have met several criteria expected for potential antifungal drug targets.

In accordance with the present invention, disruption of *CaKRE5, CaCDC24* and *CaALR1* was performed. The disruption plasmids were digested and transformed into *C. albicans* strain CA1. Southern blot analysis confirmed that the aforementioned genes are essential in *C. albicans*.

According to another aspect of the present invention, CaKRE5, CaCDC24 and CaALR1 were used in antifungal screening assays which confirmed their potential to screen for novel antifungal compounds.

While US Patent 5,194,600 claims the use of the *S. cerevisiae KRE5* gene. A number of observations from fungal biology

make it far from obvious as to the presence or role of such a gene in a pathogenic yeast, and whether it would be essential or otherwise have utility as an antifungal target. These observations are listed below.

a) A related gene, *GPT1*, in the yeast *S. pombe* is not essential, is thought to be involved in protein folding, fails to complement the *S.cerevisiae kre5* mutant, and fails to reduce b-(1,6)-glucan polymer levels in this yeast.

b) The b-(1,6)-glucan polymer could be made in a different way in different yeasts.

c) Genes are lost during evolution and it was not obvious that *C. albicans* retained a *KRE5* related gene. For example, the *CaKRE5* fails to complement a *S. cerevisiae kre5* mutant, thus no gene could be recovered by such an approach, similarly the DNA sequence of the *C. albicans CaKRE5* gene is sufficiently different from that of *S.cerevisiae*, that it cannot be detected by low stringency Southern hybridization with the *S. cerevisiae KRE5* gene as a probe.

BRIEF DESCRIPTION OF THE DRAWINGS

Having thus generally described the invention, reference will now be made to the accompanying drawings, showing by way of illustration a preferred embodiment thereof, and in which:

Figure 1 shows CaKRE5 sequence and comparison to the *S. cerevisiae KRE5*, *Drosophila melanogaster UGGT1*, and *S. pombe GPT1* encoded proteins. (A) illustrates nucleotide and predicted amino acid sequence of *CaKRE5*. The CaKRE5 signal peptide is underlined in bold. The ER retention sequence His-Asp-Glu-Leu (HDEL) is indicated in bold at the C-terminus. Non-canonical CTG codons encoding Ser in place of Leu are italicized. (B) shows protein

sequence alignment between CaKre5p, Kre5p, Gpt1p, and Uggtp. Proteins are shown in single-letter amino acid code with amino acid identities shaded in black and similarities shaded in gray. Gaps introduced to improve alignment are indicated by dashes and amino acid positions are shown at the left;

Figure 2 shows *CaALR1* sequence and comparison to *S. cerevisiae* Alr1p and Alr2p. (A) illustrates nucleotide and predicted amino acid sequence of *CaALR1*. Two hydrophobic amino acid stretches predicted to serve as transmembrane domains are indicated in bold. Non-canonical CTG codons are italicized. (B) shows protein sequence alignment between CaAlr1p, Alr1p, and Alr2p. Proteins are shown in single-letter amino acid code with amino acid identities shaded in black and similarities shaded in gray. Dashes indicate gaps introduced to improve alignment;

Figure 3 shows *CaCDC24* sequence and comparison to *CDC24* from *S. cerevisiae* and *S. pombe*. (A) illustrates nucleotide and predicted amino acid sequence of *CaCDC24*. Non-canonical CTG codons are italicized. (B) shows protein sequence alignment between CaCdc24p, *S. cerevisiae* Cdc24p, and the *S. pombe* homolog, Scd1p. The CaCdc24p dbl homology domain extends from amino acids 280-500. A pleckstrin homology domain is detected from residues 500-700. Protein alignments are formated as described in Fig. 1 and 2; and

Figure 4 illustrates disruption of *CaKRE5, CaALR1*, and *CaCDC24*. Restriction maps of (A) *CaKRE5*, (B) *CaALR1*, and (C) *CaCDC24* display restriction sites pertinent to disruption strategies. The insertion position of the *hisG-URA3-hisG* disruption module relative the *CaKRE5, CaALR1*, and *CaCDC24* open reading frames (indicated by open arrows) is indicated as well as probes used to verify disruptions by Southern blot analysis. (D-F.) show southern blot verification of targeted

integration of the *hisG-URA3-hisG* disruption module into *CaKRE5*, *CaALR1*, and *CaCDC24* and its precise excision after 5-FOA treatment. (D) shows genomic DNA extracted from *Candida albicans* wild-type strain, CAI-4 (lane 1), heterozygote *CaKRE5/cakre5Δ::hisG-URA3-hisG* (lane 2), heterozygote *CaKRE5/cakre5Δ::hisG* after 5-FOA treatment (lane 3), and a representive transformant resulting from the second round of transformation into a *CaKRE5/cakre5Δ::hisG* heterozygote (lane 4), were digested with HindIII and analyzed using *CaKRE5, hisG*, and *CaURA3* probes. Asterisks identify the 1.6 kb ladder fragment that nonspecifically hybridizes to the three probes. (E) shows genomic DNA extracted from CAI-4 (lane 1), heterozygote *CaALR1/caalr1Δ::hisG-URA3-hisG* (lane 2), heterozygote *CaALR1/caalr1Δ::hisG* after 5-FOA treatment (lane 3), and a representive transformant resulting from the second round of transformation into a *CaALR1/caalr1Δ::hisG* heterozygote (lane 4), were digested with EcoRI and analyzed using *CaALR1, hisG*, and *CaURA3* probes. (F) shows genomic DNA extracted from CAI-4 (lane 1), heterozygote *CaCDC24/cacdc24Δ::hisG-URA3-hisG* containing the disruption module in orientation 1 (lane 2), heterozygote *CaCDC24/cacdc24Δ::hisG-URA3-hisG* containing the disruption module in orientation 2 (lane 3), heterozygote CaALR1/caalr1D::hisG (orientation 1) after 5-FOA treatment (lane 4), heterozygote *CaALR1/caalr1Δ::hisG* (orientation 2) after 5-FOA treatment (lane 5) and a representive transformant resulting from the second round of transformation into a *CaALR1/caalr1Δ::hisG* (orientation 1) heterozygote (lane 6), were digested with EcoRI and analyzed using *CaCDC24, hisG*, and *CaURA3* probes.

Other objects, advantages and features of the present invention will become more apparent upon reading of the following

non-restrictive description of preferred embodiments with reference to the accompanying drawing which is exemplary and should not be interpreted as limiting the scope of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

We have identified *C. albicans* genes homologous to the essential genes *KRE5, CDC24*, and *ALR1* from *S. cerevisiae*. These genes participate in essential cellular functions of cell wall biosythesis, polarized growth, and divalent cation transport, respectively. Disruption of these genes in *C. albicans* experimentally demonstrates their essential role in this pathogenic yeast. Database searches fail to identify clear homologous counterparts in mammalian genomes, supporting the utility of these genes as novel antifungal targets.

KRE5

The *S. cerevisiae KRE5* gene meets several criteria expected for a potential antifungal drug target. Deletion of *KRE5* confers a lethal phenotype (2). Although *KRE5*-deleted cells are known to be viable in one particular strain background, they are extremely slow growing and spontaneous extragenic suppressors are required to propagate *kre5D* cells under laboratory conditions. Genetic analyses suggest that *KRE5*, together with a number of additional *KRE* genes participates in the *in vivo* synthesis of β-(1,6)-glucan. β-(1,6)-glucan covalently cross-links or "glues" other cell surface constituents, namely β-(1,3)-glucan, mannan, and chitin into the final wall structure and and has been shown to be essential for viability in both *S. cerevisiae* and *C. albicans* (1,2 and references therein). Moreover, β-(1,6)-glucan has been demonstrated to exist in a number of additional fungal classes including other yeast and filamentous *Ascomycetes*, *Basidiomycetes* and

Oomycetes. Importantly, however, efforts have failed to detect β-(1,6)-glucan in higher eukaryotes.

Consistent with a role in β-(1,6)-glucan biosynthesis, *in vivo* levels of this polymer are reduced substantially in *kre5-1* cells versus an isogenic wild type strain, and are completely absent in several independently-suppressed *kre5* null strains (2). In addition, *kre5* mutants show a number of genetic interactions with *kre6*, another gene involved in β-(1,6)-glucan assembly [Shahinian and Bussey, personal communication)]. Although the biochemistry of β-(1,6)-glucan synthesis remains poorly understood, recent studies demonstrate that cell wall mannoproteins are extensively glucosylated through β-(1,6) linkages and that this modification plays a central role in their anchorage within the extracellular matrix. KRE5 plays a critical role in this process as well, as Cwp1p, an abundent cell wall protein which is demonstrated to be highly glucosylated through β-(1,6)-glucan addition, is undetected in the cell wall fraction of *kre5D* cells, and instead secreted into the medium.

The predicted *KRE5* gene product offers only limited insight into a possible biochemical activity related to β-(1,6)-glucan production. *KRE5* encodes a large secretory protein containing both an N-terminal signal peptide and C-terminal HDEL retention signal for localization to the endoplasmic reticulum. Interestingly, Kre5p has limited but significant homology to UDP-glucose:glycoprotein glycosyltransferases (UGGT), an enzyme class participating in the "quality control" of protein folding. Such UGGT enzymes function to "flag" misfolded ER proteins by reglucosylation of N-linked GlcNAc2Man9 core oligosaccharide structures present on misfolded proteins. Proteins labelled in this way are substrates for the ER chaperonin, calnexin, which facilitates refolding of the misfolded protein. However, genetic

analyses to address the relative involvement of KRE5 in glucosylation-dependent protein folding and β-(1,6)-glucan biosynthesis demonstrate that the essential function of KRE5 is unrelated to protein folding, and instead relates to its role in β-(1,6)-glucan polymer biosynthesis (3). Although it remains to be demonstrated biochemically, KRE5 homology to glycosyltransferases likely reflects its role in the early biosynthesis of this polymer.

ALR1

The product of the *S. cerevisiae* gene, *ALR1*, also meets several of the conditions necessary for a suitable antifungal drug target. Strains deleted of *ALR1* show limited growth with supplementary Mg^{+2} but and are otherwise inviable (4). These results demonstrate that *ALR1* is essential for growth. *ALR1* encodes a 922 amino acid protein containing a highly charged N-terminal domain and two hydrophobic C-terminal regions predicted to serve as membrane spanning domains anchoring the protein at the plasma membrane. Although such a localization remains to be directly demonstrated, deposition to the cell surface makes Alr1p an attractive drug target in terms of both bioavailability and resistance issues (see Discussion). Alr1p shares substantial homology to two additional *S. cerevisiae* proteins, Alr2p (70% identity) and Ykl064p (34% identity). Both Alr1p and Alr2p share limited similarity to CorA, a Salmonella typhimurium protein periplasmic membrane protein involved in divalent cation transport. Mammalian homologues to *ALR1* have not been detected despite extensive database searches and the gene is absent from the metazoan *Caenorhabditas elegans*.

Although ALR1 was identified in a screen for genes that confer increased tolerance to Al^{+3} when overexpressed, biochemical

analyses support a role for *ALR1* in the uptake system for Mg^{+2} and possibily other divalant cations. Mg^{+2} is an essential requirement for bacterial and yeast growth. Uptake of radiolabelled Co^{+2}, an analog of Mg^{+2} for uptake assays, correlates with *ALR1* activity. Overexpression of *ALR1* increased Co^{+2} uptake four-fold, while deletion of ALR1 substantially reduced uptake. As mentioned above, Alr1p shares structural and sequence similarity to CorA, an extensively characterized Mg^{+2} import protein and deletion of *ALR1* is only suppressed with the addition of supplementary Mg^{+2}.

CDC24

A third potential antifungal drug target is the *S. cerevisiae* gene, *CDC24*. Accordingly, *CDC24* is essential for viability in both *S. cerevisiae* and *S. pombe* (5). *CDC24* has been biochemically demonstrated to encode GDP-GTP nucleotide exchange factor (GEF) activity towards Cdc42p, a Rac/Rho-type GTPase involved in polarization of the actin cytoskeleton. Conditional alleles of *CDC24* shifted to the nonpermisive temperature lack a polarized distibution of actin, and consequentially form large, spherical, unbudded cells in which the normal polarized deposition of cell wall material is disrupted. Eventually *cdc24* mutants lyse at the restrictive temperature. *CDC24*-dependent activation of *CDC42*, is also required for the activation of the pheromone response signal transduction pathway during mating, and likely participates in the activation of this pathway under conditions that promote pseudohyphal development, since a downstream effector of *CDC42*, *STE20*, is required for hyphal formation. Thus *CDC24* regulates cell wall assembly and the yeast-hyphal dimorphic transition; both key cellular processes and targets being actively pursued in antifungal drug screens.

Cdc24p localizes to the cell cortex concentrating at sites of polarized growth and interacts physically with a number of proteins including Cdc42p, Bem1p, and the heterotrimeric G protein β and γ subunits encoded by *STE4* and *STE18* respectively. Cdc24p shares 24% overall identity to its *S. pombe* counterpart, Scd1p. Similar homology has not been found in mammalian database protein searches, although Cdc24p does possess limited homology to a domain of the human exchange protein, dbl, and contains a pleckstrin homology domain, common to several mammalian protein classes. Unlike this limited homology to Cdc24p outside of fungi, Cdc42p conversely shares 80-85% identity to mammalian isoforms. Perhaps the fungal-specificity of *CDC24* may be due to its role in the fungal-specific processes of bud formation, pseudohyphal growth, and projection formation during mating, whereas *CDC42* performs highly conserved functions (namely actin polymerization and signal transduction) common to all eukaryotes.

Isolation of *CaKRE5, CaCDC24,* and *CaALR1.*

To isolate full length clones of CaKRE5, CaCDC24, and CaALR1, oligonucleotides were designed according to publicly available fragments of C. albicans DNA sequence. Polymerase chain reaction (PCR) using oligonucleotide pairs CAKRE5.1/CAKRE5.2, CaCDC24.1/CaCDC24.2, and CaALR1.1/CaALR1.2 to amplify genomic DNA derived from C. albicans strain SC5314 yielded 574, 299, and 379 bp products, respectively. These PCR products were ^{32}P-radiolabeled and used to probe a YEp352-based *C. albicans* genomic library by colony hybridization.

Sequence Information

DNA sequencing of two independent isolates representing putative *CaKRE5* and *CaALR1* clones revealed complete open reading frames sharing statistically significant homology to their *S. cerevisiae* counterparts (Fig. 1, 2). DNA sequencing of multiple isolates of *CaCDC24* revealed an orf containing strong identity to *CDC24*, but predicted to be truncated at its 3' end. The 3' end of *CaCDC24* was isolated by PCR amplification using one oligonucleotide designed from its most 3' sequence and a second oligonucleotide which anneals to the YEp352 polylinker allowing amplification of *CaCDC24* C-terminal encoding fragments from this *C. albicans* genomic library. Subcloning and DNA sequencing of a 1.0 kb PCR product completes the *CaCDC24* open reading frame and reveals its gene product to share strong homology to both Cdc24p and Scd1p (Fig. 3).

CaKRE5

Sequence analysis reveals *CaKRE5* and *KRE5* are predicted to encode similarly-sized proteins (1447 vs 1365 amino acids; 166 vs 156 kDA) sharing significant homology throughout their predicted protein sequences (22% identity, 42% similarity, (Fig. 1)). Moreover, like *KRE5*, *CaKRE5* is predicted to possess an amino-terminal signal peptide required for translocation into the secretory pathway, and a C-terminal HDEL sequence which facilitates the retention of soluble secretory proteins within the endoplasmic reticulum (ER). Although CaKre5p is more homologous to *S.pombe* and metazoan UGGT proteins throughout its C-terminal domain than to Kre5p, CaKre5p and Kre5p, they are more related to each other over their remaining sequence (approx. 1100 amino acids). This unique homology between the two proteins as well as a similar null phenotypes (see

below) suggest that *CaKRE5* likely serves as the *KRE5* counterpart in *C. albicans*.

CaALR1

CaALR1 encodes a 922 amino acid residue protein sharing strong identity to both ALR1 (1.0e-180) and ALR2 (1.0e-179, (Fig.2)). Like these proteins, CaALR1 possesses a C-terminal hydrophobic region which likely functions as two transmembrane anchoring domains ( ). CaALR1 shares only limited homology, however, to two highly homologous regions common to ALR1 and ALR2; neither the N-terminal 250 amino acids of CaALR1 nor its last 50 amino acids C-terminal the hydrophobic domain share strong similarity to ALR1 or ALR2. In addition, CaALR1 possesses two unique sequence extentions within the CorA homology region (one 38 a.a. in length, the other, 16 a.a. long) not found in either ALR1 or ALR2. Protein database searches identify a *S.pombe* hypothetical protein sharing strong homology to CaALR1 (2.7e-107), however no similarity to higher eukaryotic proteins were detected.

CaCDC24

Sequence analysis of the *CaCDC24* gene product reveals extensive homology to both Cdc24p (3.8e-97) and Scd1p (1.0e-59, Fig.3)) throughout their entire open reading frames. Although substantial similarity exists between CaCdc24p (and both Cdc24p and Scd1p) and a large number of metazoan proteins (upto 1.8e-13), in each case this homology is restricted to either the nucleotide exchange domain, (dbl domain), or a domain common to signal transduction components (PH domain). Extensive database searches reveal that both the N-terminal (250 a.a.) and C-terminal (300 a.a.) regions of

CaCdc24p are exclusively conserved within this fungal family of homologs.

Disruption of CaKRE5, CaALR1, and CaCDC24

Experimental strategy

Disruption of *CaKRE5* was performed using the *hisG-CaURA3-hisG* "*URA*-blaster" cassette constructed by Fonzi and Irwin and standard molecular biology techniques (1, and references within). A *cakre5::hisG-CaURA3-hisG* disruption plasmid was constructed by deleting a 780bp BamH1-BglII DNA fragment from the library plasmid isolate, p*CaKRE5*, and replacing it with a 4.0 kb BamHI-BglII DNA fragment containing the *hisG-CaURA3-hisG* module from pCUB-6. This *CaKRE5* disruption plasmid is deleted of DNA sequence encoding amino acids 971-1231, which encompasses approx. 50% of the UGGT homology domain. This *CaKRE5* disruption plasmid was then digested with SphI prior to transformation.

A *CaALR1* disruption allele was constructed by first subcloning a 7.0 kp *CaALR1* BamHI-SalI fragment from YEp352-library isolate p*CaALR1* into PBSKII+. A 841 bp *CaALR1* HindIII-BglII fragment was then replaced with a 4.0 kb *hisG-CaURA3-hisG* DNA fragment digested with HindIII and BamHI from PBSK-*hisG-CaURA3-hisG*. This *CaALR1* disruption allele, which is lacking DNA sequences encoding amino acids 20-299, was digested using BamHI and SalI prior to transformation.

A *CaCDC24* insertion allele was constructed by first deleting a 0.9 kb KpnI fragment from YEp352-library isolate p*CaCDC24* to remove *CaCDC24* upstream sequence containing BamHI and BglII restriction sites which obstruct the insertion of the *hisG-CaURA3-hisG* module. The 4.0 kb BamHI-BglII *hisG-CaURA3-hisG* fragment from

pCUB-6 was then ligated into a unique BglII site in p*CaCDC24*-Kpn1D. The resulting plasmid, p *cacdc24::hisG-CaURA3-hisG*, possessing an insertion allele within *CaCDC24* at amino acid position 306, was digested with KpnI and SalI prior to transformation.

CaKRE5, CaALR1, and *CaCDC24* disruption plasmids were digested as described above, and transformed into *C. albicans* strain CAI4 using the lithium acetate method. Transformants were selected as Ura+ prototrophs on YNB + Casa plates. Heterozygous disruptants were identified by PCR (data not shown), verified by Southern blot (see below), and prepared for a second round of gene disruption by selecting for 5-FOA resistance. To assess the null phenotype of each gene, a second round of transformations using heterozygous *CaKRE5/cakre5*, *CaALR1/caalr1*, and *CaCDC24/cacdc24 ura3-* strains were performed as outlined above.

Correct integration of the *hisG-CaURA3-hisG* module into *CaKRE5*, *CaALR1*, and *CaCDC24* and *CaURA3* excision from heterozygous strains were verified by Southern blot analysis using the following probes:

(1a) a 1.25 kb XbaI-Kpn1 fragment digested from p*CaKRE5* containing N-terminal coding sequence of *CaKRE5*;

(1b) a 1.7 kb PCR product containing coding sequence from amino acid 404 and 3' flanking sequences of *CaALR1*;

(1c) a 778 bp PCR product containing *CaCDC24* coding sequence from amino acids 154-430;

(2) a 783 bp PCR product which contains the entire *CaURA3* coding region;

(3) a 898bp PCR product encompassing the entire *Salmonella typhimurium hisG* gene. Genomic DNA from *CaKRE5*-disrupted strains were digested with HindIII and EcoR1 was

used to digest genomic DNA from *CaALR1* and *CaCDC24*-disrupted strains.

Results

Southern blot analysis revealed that the *cakre5::hisG-CaURA3-hisG* disruption fragment integrated precisely into the wild type locus (Fig 4D) after the first round of transformations. Both a 5.0 kb wild type band and a 9.0 kb band diagnostic of the *CaKRE5*-disrupted allele were detected using the *CaKRE5* probe (Fig 4D). The 9.0 kb band was also detected with both the *hisG* and *CaURA3* probes, confirming disruption of the first *CaKRE5* copy. Successful excision of the *CaURA3* gene by growth on 5-FOA was validated by 1) a predicted shift in size of the *CaKRE5* disruption fragment from 9.0 kb to 6.0 kb when probed with either *CaKRE5* or *hisG* probes and 2) the inability of the CaURA3 probe to recognize this fragment and the resulting strain having reverted to ura3- prototrophy.

To determine whether *CaKRE5* is essential, the transformation was repeated in two independently-derived *CaKRE5/cakre5::hisG, ura3-/ura3-* heterozygous strains. A total of 36 Ura+ colonies (24 small and 12 large colonies after 3 days of growth) were analyzed by PCR using oligonucleotides which amplify a 2.5 kb wild-type fragment that spans the BamHI and BglII sites bordering the disrupted region. All colonies were determined to contain this 2.5 kb wild-type fragment but lacking the 2.8 kb *cakre5::hisG* allele, consistent with the *cakre5::hisG-CaURA3-hisG* module integrating at the disrupted locus. Southern blot analysis using the 3 different probes independently confirmed 4 such Ura+ transformants as *bonafide CaKRE5/cakre5::hisG-CaURA3-hisG* heterozygotes. If disruption of both copies of the gene were not essential then 50% of the recovered

disruptants are expected to integrate into the *CaKRE5* locus giving homologous disruptants and 50% being heterozygous. For example, this is the case when disrupting the second wild-type allele of *CaKRE1*; a gene shown not to be essential in *S. cerevisiae*. An equal number of heterozygous and homozygous strains result from this second round of transformations (data not shown). However, the absence of any homozygous *CaKRE5* disrupted transformants being detected among the 36 Ura+ transformants analyzed supports our contention that *CaKRE5* is essential in *C. albicans*.

CaALR1

Southen blot analysis of *CaALR1* first round transformants confirmed correct integration of the *caalr1::hisG-CaURA3-hisG* disruption module as judged by an appropriately sized disruption band of 5.7 kb, and a wild-type fragment predicted to be >9.0 kb detected by the *CaALR1* probe (Fig 4E). This 5.7 kb band was also detected with both the *hisG* and *CaURA3* probes, confirming disruption of one copy of *CaALR1*. Southern blotting confirmed excision of the *CaURA3* gene by growth on 5-FOA as the *CaALR1* probe detected an expected 5.0 kb fragment due to the absense of *CaURA3*. Moreover, this 5 kb *caalr::hisG* band was also detected using the *hisG* probe but not with the *CaURA3* probe (Fig. 4E).

Determination of the *CaALR1* null phenotype was performed as described for *CaKRE5*. However, as it has been reported that the inviability of the *ALR1* null mutation in *S. cerevisiae* can be partially suppressed by supplementing the medium with $MgCl_2$, we performed the second transformation by selecting for Ura+ colonies on 500mM $MgCl_2$-containing medium as well or standard Casa plates.

35+ colonies of various size (22 from $MgCl_2$ -supplemented plates) were analyzed by PCR to confirm *caalr1::hisG-CaURA3-hisG* integration. The second allele from each of these 35 transformants was determined to be wild-type by PCR using oligos that span the insertion and produce a wild-type 1.6 kb product and not the slightly larger 1.75 kb product of the *caalr::hisG* allele (Note. this was done 2X/run far in 2% agarose/and alongside *Caalr::hisG* control genomic DNA which did run noticably slower than the 35 unknowns). Southern blot analysis using the 3 different probes independently confirmed 4 such Ura+ transformants as *CaALR1/caalr1::hisG-CaURA3-hisG* heterozygotes. Our inability to identify a homozygous CaALR1 disrupted transformant among the 35 Ura+ colonies analyzed, supports the claim that *CaALR1* is essential in *C. albicans*.

CaCDC24

Southern blot analysis of *CaCDC24* first round transformants using the *CaCDC24* gene probe confirmed correct integration of the *cacdc24::hisG-CaURA3-hisG* insertion fragment as both 2.55 kb and 3.7 kb fragments, diagnostic of the insertional allele, were detected in addition to the 2.2 kb wild-type CaCDC24 fragment (Fig. 4F). Moreover, both 2.55 kb and 3.7 kb fragments were detected using *CaURA3* and *hisG* probes. Excision of *CaURA3* from the resulting heterozygote was verified by 1) detecting a single 3.3 kb fragment unique to 5-FOA resistant colonies using ther *CaCDC24* or *hisG* probes, and 2) the failure to detect this band using the *CaURA3* probe. (Fig. 4F).

A second round of transformations using the above described *CaCDC24* heterozygote was performed. 28+ colonies of various size were analyzed by PCR to confirm *cacdc24::hisG-CaURA3-hisG* integration. The second allele from each

of these 28 transformants was determined to be wild-type by PCR using oligos that span the insertion and produce a wild-type 0.5 kb product and not the 1.6 kb product of the *caalr::hisG* allele. Southern blot analysis using the 3 different probes independently confirmed 4 such Ura+ transformants as *CaCDC24/cacdc24::hisG-CaURA3-hisG* heterzygotes. Our inability to identify a homozygous *CaCDC24* disrupted transformant among these 28 Ura+ colonies analyzed, strongly suggests that *CaCDC24* is essential in *C. albicans* like it is known to be in *S. cerevisiae*.

The present invention is illustrated in further detail by the following non-limiting examples.

EXAMPLE 1

In vivo Screening Methods for Specific Antifungal Agents

Candida albicans strains with reduced or elevated levels of the *CaKRE5, CaALR1*, or *CaCDC24* gene product permit screens for differential sensitivity or resistance to a drug or compounds from natural or artificial sources that inhibit these proteins. Compounds that show such a differential inhibition of growth of such *Candida albicans* strains would be specific inhibitors of *CaKRE5, CaALR1*, or *CaCDC24*-dependent processes and can be further evaluated as specific antifungal drugs.

Expression of a functional *CaKRE5, CaALR1*, or *CaCDC24* in a *S.cerevisiae kre5, alr1* and *cdc24* mutant respectively, allows replacement of the *S. cerevisiae* gene with that of its *C. albicans* counterpart and thus permits screening for specific inhibitors in a *S. cerevisiae* background where the additional experimental tractability of the organism permits additional sophistication of the screens. For example, drugs which block CaKre5p in *S. cerevisiae* confer

K1 killer toxin resistance, and this phenotype can be used to screen for such compounds. Similarly, drugs/compounds could be screened which inactivate heterologously-expressed *CaCDC24* and consequently disrupt its association with Rsr1p or Cdc42p in a two hybrid assay. Alternatively, *CaCDC24* function could be monitored in a screen for compounds able to disrupt pseudohyphal formation in a *CaCDC24*-dependent manner. A whole cell drug screening assay based on *CaALR1* function could similary be envisaged. For example, *CaALR1*-dependent influx of 57_{CO2+} in a *S. cerevisiae alr1* mutant suppressed by supplementary Mg^{2+} could be monitored to identify compounds which specifically block the import of divalent cations.

EXAMPLE II

In vitro Screening Methods for Specific Antifungal Agents

1. Use of an *in vitro* assay to synthesize β-(1,6)-glucan.

In such an assay the incorporation of labelled glucose from UDP-glucose into a product that can be immunoprecipitated or immobilized with β-(1,6)-glucan antibodies is measured. The specificity of this synthesis can be established by showing its dependence on CaKre5p, and its digestion with β-(1,6)-glucanase.

Drugs which block this *in vitro* synthesis reaction, block β-(1,6)-glucan synthesis and are candidates for antifungal drugs, some may inhibit Kre5p, others may inhibit other steps in the synthesis of this polymer.

2. Use of a specific in vitro assay for CaKre5p.

CaKre5p has amino-acid sequence similarities to UDP-glucose glycoprotein glucosyltransferases. The CaKre5p protein can be produced heterogeneously or from *Candida albicans* and an

assay devised using a range of substrates that are subset of glycoproteins that are in the wall with GPI modifications that are β-(1,6)-glucosylated. These acceptor substrates would be obtained from a strain of *S. cerevisiae* that is a *kre5* disruption and have failed to receive the glucose from the UDP-glucose donor to the acceptor substrate *in vivo*. Such an assay measuring CaKre5p dependent protein glycosylation can be used to screen for inhibitors of CaKre5p. Alternatively, it would be possible to screen for compounds that bind to immobilised CaKre5p. Such inhibitors and Kre5p-binding proteins would be candidates for drugs specifically inhibiting this fungal-specific process.

CDC24 has been biochemically demonstrated to encode a GDP-GTP nucleotide exchange factor (GEF) required to convert Cdc42p to a GTP-bound state. An *in vitro* assay to measure CaCdc24p-dependent activation of Cdc42p could be used to screen for inhibitors of CaCDC24p. This could be accomplished by directly measuring the percentage of GTP versus GDP bound by Cdc42p. Alternatively, Cdc24p function could be determined indirectly by measuring Cdc42p-GTP dependent activation of Ste20p kinase activity.

EXAMPLE III

The use of CaALR1, CaKRE5, and CaCDC24 in PCR-based diagnosis of fungal infection

Polymerase chain reaction (PCR) based assays provide a number of advantages over traditional serological testing methodologies in diagnosing fungal infection. Issues of epidemiology, fungal resistance, relability, sensitivity, speed, and strain identification are limited by the spectrum of primers and probes available. The

CaKRE5, *CaALR1*, and *CaCDC24* gene sequences enable the design of novel primers of potential clinical use. In addition, as CaALR1 is thought to localize to the plasma membrane and extend out into the periplasmic space/cell wall, this extracellular domain could act as a serological antigen to which antibodies could be raised and used in serological diagnostic assays.

EXAMPLE IV

Plasmid-based reporter constructs which measure Kre5p, Alr1p, or Cdc24p inactivation

Transcriptional profiling of *kre5*, *alr1*, and *cdc24* mutants in *S. cerevisiae* to identify genes which are transcriptionally induced/repressed specifically under conditions of *KRE5*, *ALR1*, or *CDC24* inactivation or overproduction. The identification of promoter elements from genes responsive to the loss of *KRE5*, *ALR1*, or *CDC24* activity offers practical utility in drug screening assays to identify compounds which specifically inactivate these targets. For example, a chimeric reporter gene (eg. *lacZ*, *GFP*,) whose expression would be induced/repressed by such a promoter would reflect activity of Kre5p, and could be used for high-throughput screening of compound libraries. Further a group of promoters showing such regulated expression would allow a specific fingerprint or transcriptional profile to be built for the inhibition or overproduction of the *ALR1*, *CDC24*, or *KRE5* genes. This would allow a reporter set to be constructed that could be used for high-throughput screening of compound libraries giving a specific tool for screening compounds which inhibit these gene products.

CONCLUSION

We have identified the *CaKRE5, CaALR1*, and *CaCDC24* genes from *C. albicans* and validated their utility as novel antifungal drug targets by demonstrating their essential nature by gene disruption. Although the precise function of their gene products remains to be determined, we have shown that these proteins are essential for viability. Genome database searches fail to detect significant homology to these genes in metazoans, suggesting that screening for compounds which inactivate these fungal-specific drug targets are less likely to display toxicity to human cells. *KRE5* and *CDC24* are unique genes in *S. cerevisiae* and irrespective of being members of gene families in *C.albicans*, they retain an essential function. *Alr1p1* is part of a 3 member gene family in *S. cerevisiae*, and sequence similarity to Alr2p has been identified (Stanford Sequencing Project), however the essential role of CaALR1p in *C. albicans* and their predicted extracellular location offers the potential to screen for novel antifungal compounds which need not enter the cell, circumventing issues of compound delivery and drug resistance.

We have shown that the *Candida albicans CaKRE5* gene is essential; has a protein product with significant sequence similarity to *S. cerevisiae* Kre5p at the gene product level, and is involved in β-(1,6)-glucan synthesis as there is a reduced amount of the polymer in a heterozygous *CaKRE5/Cakre5* disruption relative to the *CaKRE5/CaKRE5* homozygote, and the phenotype of the heterozygous *CaKRE5/Cakre5* disruption mutant cells resembles that of *kre5* deletions in *S.cerevisiae*, clumps of swollen cells with cytokinesis and cell separation defects (data not shown).

Thus, in the present invention we reduce to practice the use of *CaKRE5, CaALR1*, and *CaCDC24* in *Candida albicans* as

essential antifungal targets, and extend in a non-obvious way the use of these genes to a pathogenic fungal species as targets for screening for drugs specifically directed against fungal pathogens.

Although the present invention has been described hereinabove by way of preferred embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims.

REFERENCES

1. Lussier, M., A-M Sdicu, S. Shahinian, and H. Bussey. 1998. The Candida albicans KRE9 gene is required for cell wall b-(1,6)-glucan synthesis and is essential for growth on glucose. Proc. Natl. Acad. Sci. USA 95:9825-9830.

2. Meaden, P., K. Hill, J. Wagner, D. Slipetz, S. S. Sommer, and H. Bussey. 1990. The yeast KRE5 gene encodes a probable endoplasmic reticulum protein required for b-(1,6)-glucan synthesis and normal cell growth. Mol. Cell. Biol. 10:3013-3019.

3. Orlean, P. 1997. In The Molecular and Cellular Biology of the Yeast Saccharomyces, eds. Pringle, J. R., Broach, J. R., and Jones, E. W. Cold Spring Harbor Lab. Press, Plainview, NY. Vol 3, pp229-362.

4. Shahinian, S., G. J. P. Dijkgraaf, A-M Sdicu, D. Y. Thomas, C. A. Jakob, M. Aebi, and H. Bussey. 1998. Involvement of protein N-glycosyl chain glucosylation and processing in the biosynthesis of cell wall b-(1,6)-glucan of Saccharomyces cerevisiae. Genetics 149:843-856.

5. MacDiarmid, C. W., and R. C. Gardner. 1998. Overexpression of the Saccharomyces cerevisiae magnesium transport system confers resistance to aluminum ion. J. Biol. Chem. 273:1727-1732.

6. Pringle, J et al. 1995. Establishment of cell polarity in yeast. Cold Spring Harbor Symp. Quant. Biol. 60: 729-744.

WHAT IS CLAIMED IS:

1. An isolated DNA sequence selected from the group consisting of:

a) fungal specific gene of *C. albicans* termed *CaKRE5;*

b) fungal specific gene of *C. albicans termed CaALR1;*

c) fungal specific gene of *C. albicans* termed CaCDC24;

d) a part or oligonucleotide derived from a), b) or c);

e) a nucleotide sequence complementary to any of the nucleotide sequences of a) - d); and

f) a sequence which hybridizes under high stringency conditions to any of the nucleotide sequences of a) - e).

2. The isolated DNA sequence of claim 1, wherein said sequence of *CaKRE5* is as set forth in Figure 1A.

3. The isolated DNA sequence of claim 1, wherein said sequence of *CaALR1* is as set forth in Figure 2A.

4. The isolated DNA sequence of claim 1, wherein said sequence of *CaCDC24* is as set forth in Figure 3A.

5. A method of selecting a drug that modulates the activity of a protein encoded by said *CaKRE5* of claim 2 comprising:

a) incubating a candidate drug with said protein;

b) determining the activity of said protein in the presence of said candidate drug,

wherein a potential drug is selected when the activity of said protein in the presence of said candidate drug is measurably different than in the absence thereof.

6. A method of selecting a drug that modulates the activity of a protein encoded by said *CaALR1* of claim 3 comprising:

a) incubating a candidate drug with said protein;

b) determining the activity of said protein in the presence of said candidate drug,

wherein a potential drug is selected when the activity of said protein in the presence of said candidate drug is measurably different than in the absence thereof.

7. A method of selecting a drug that modulates the activity of a protein encoded by said CaCDC24 of claim 3 comprising:

a) incubating a candidate drug with said protein;

b) determining the activity of said protein in the presence of said candidate drug,

wherein a potential drug is selected when the activity of said protein in the presence of said candidate drug is measurably different than in the absence thereof.

8. An isolated nucleic acid molecule consisting of 10 to 50 nucleotides which specifically hybridizes to RNA or DNA of claim 1, 2, 3 or 4, wherein said nucleic acid molecule is or is complementary to a nucleotide sequence consisting of at least 10 consecutive nucleotides from said nucleic acid sequence set forth in Figures 1A, 2A or 3A.

9. A method of detecting *CaKRE5*, *CaALR1* or *CaCDC24* in a sample comprising:

a) contacting said sample with a nucleic acid molecule according to claim 8, under conditions such that hybridization occurs; and

b) detecting the presence of said molecule bound to said *CaKRE5, CaALR1* or *CaCDC24* nucleic acid.

10. A purified CaKRE5 polypeptide or an epitope-bearing portion thereof.

11. A purified CaALR1 polypeptide or an epitope-bearing portion thereof.

12. A purified CaCDC24 polypeptide or an epitope-bearing portion thereof.

13. The purified CaKRE5 polypeptide according to claim 10, comprising an amino acid sequence at least 90% identical to the amino acid sequence as set forth in Figure 1B.

14. The purified CaALR1 polypeptide according to claim 11, comprising an amino acid sequence at least 90% identical to the amino acid sequence as set forth in Figure 2B.

15. The purified CaCDC24 polypeptide according to claim 12, comprising an amino acid sequence at least 90% identical to the amino acid sequence as set forth in Figure 3B.

16. An antibody having specific binding affinity to the polypeptide or epitope-bearing portion thereof according to claim 10.

5

ABSTRACT OF THE DISCLOSURE

The invention relates to the identification and disruption of essential fungal specific genes isolated in the yeast pathogen *Candida albicans* namely *CaKRE5, CaALR1* and *CaCDC24* and to the use thereof in antifungal diagnosis and as essential antifungal targets in a fungal species for antifungal drug discovery.

VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS (37 CFR 1.9(f) AND 1.27 (d)) - NONPROFIT ORGANIZATION	Docket No 11168.93

Serial No.	Filing Date	Patent No.	Issue Date

Applicant/
Patentee: Terry ROEMER et al.

Invention: IDENTIFICATION OF THE CANDIDA ALBICANS ESSENTIAL FUNGAL SPECIFIC GENES CaKRE5, CaALR1 AND CaCD24 AND USE THEREOF IN ANTIFUNGAL DRUG DISCOVERY.

I hereby declare that I am an official empowered to act on behalf of the nonprofit organization identified below

NAME OF ORGANIZATION: McGILL UNIVERSITY
ADDRESS OF ORGANIZATION 845 Sherbrooke Street West
Montreal, Quebec, Canada, H3A 1B1

TYPE OF NONPROFIT ORGANIZATION

- ☒ University or other Institute of Higher Education
- ☐ Tax Exempt under Internal Revenue Service Code (26 U S C 501(a) and 501(c)(3))
- ☐ Nonprofit Scientific or Educational under Statute of State of The United States of America
  Name of State: Citation of Statute
- ☐ Would Qualify as Tax Exempt under Internal Revenue Service Code (26 U S C. 501(a) and 501(c)(3)) if Located in The United States of America
- ☐ Would Qualify as Nonprofit Scientific or Educational under Statute of State of The United States of America if Located in The United States of America
  Name of State: Citation of Statute

I hereby declare that the above-identified nonprofit organization qualifies as a nonprofit organization as defined in 37 C.F.R. 1.9(e) for purposes of paying reduced fees to the United States Patent and Trademark Office regarding the invention described in

- ☒ the specification to be filed herewith
- ☐ the application identified above
- ☐ the patent identified above.

I hereby declare that rights under contract or law have been conveyed to and remain with the nonprofit organization with regard to the above identified invention

If the rights held by the above-identified nonprofit organization are not exclusive, each individual, concern or organization having rights to the invention is listed on the next page and no rights to the invention are held by any person, other than the inventor, who could not qualify as an independent inventor under 37 CFR 1.9(c) or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e)

Each person, concern or organization to which I have assigned, granted, conveyed, or licensed or am under an obligation under contract or law to assign, grant, convey, or license any rights in the invention is listed below.

☒ no such person, concern or organization exists
☐ each such person, concern or organization is listed below

FULL NAME _____
ADDRESS _____
☐ Individual ☐ Small Business Concern ☐ Nonprofit Organization

FULL NAME _____
ADDRESS _____
☐ Individual ☐ Small Business Concern ☐ Nonprofit Organization

FULL NAME _____
ADDRESS _____
☐ Individual ☐ Small Business Concern ☐ Nonprofit Organization

FULL NAME _____
ADDRESS _____
☐ Individual ☐ Small Business Concern ☐ Nonprofit Organization

Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 CFR 1 27)

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1 28(b))

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

NAME OF PERSON SIGNING: Dr. Alex NAVARRE
TITLE IN ORGANIZATION: Director, Office of Technology Transfer
ADDRESS OF PERSON SIGNING: 3550 University Street
Montreal, Quebec, Canada, H3A 2A7

SIGNATURE: _____ DATE: 03-05-99

Figure 1A.

```
1        TGCGAATGAAAGCAAAGTTTTAGGACTGAATTCAAAAATTAATAAACCTATCGGTGGCTCGCCAAGGGATCTTAGACAGGCAATTCTAGAGGCGGCAGAGCGTCGGTTG
110 AGAGATTCAAAATGGTGTCATAGTGAAAATGCAGAAACCGAAAGTGTTCCCAAAGAGGACGAGTACGACACAACTCAGGTGGAGCTTATCGGTCCTACAGAAGGTAAACCAGTTG
225 GAACATTTGCTAATGATATCATTGATTTAACATCGGACACTGAAGAAACTCCAATTCAACCTGATAACCCGAAACGCCGCATACTCCACGAGATAATTGATTTAACTTCAGATAC
340 AGAAGACATAGAGCCAACATCACCAGAGGTAATATGTATAGATTAAGTTAAATATAAAGGCAAATATATTGCCAATGTAATACTCTTTTAACAGTGTTGTTCTCGTGCAAGGATT
455 AAGCACCGAAAAAAAAATATGTGGATGCGTTGTTATTAGTTTTACTCTTTGCTTTTTCTGAAAAGAAACATTAACGTGTTCTACTAGTTTGTCACACTACGACACAAGTCCTTGAA
    Met Ser Phe Ala Arg Tyr Ile Tyr Tyr Thr Ile Ala Val Ala Val Leu Leu Asn Phe Val Lys Ala Thr Glu Asn Asn Asn Phe Lys    29
570 ATG TCA TTT GCA AGG TAT ATC TAC TAC ACC ATT GCG GTT GCT GTT TTA TTA AAT TTT GTC AAA GCT ACT GAA AAT AAC AAT TTT AAA
    Leu Glu Val Glu Ala Ser Trp Ser Asn Ile Asp Phe Leu Pro Ser Phe Ile Glu Ala Ile Val Gly Phe Asn Asp Ser Leu Tyr Glu    58
657 CTT GAA GTT GAA GCG TCA TGG AGC AAT ATT GAT TTC CTT CCT AGC TTT ATA GAG GCC ATC GTT GGC TTC AAT GAC TCT TTG TAC GAA
    Gln Thr Ile Glu Thr Ile Phe Gly Leu Gly Asp Thr Glu Val Glu Leu Glu Asp Asp Ala Ser Asp Gln Glu Ile Tyr Ser Thr Val    87
744 CAG ACA ATT GAA ACA ATT TTT GGT TTA GGA GAC ACT GAA GTG GAA TTA GAA GAT GAT GCT TCA GAT CAA GAA ATA TAT TCT ACC GTG
    Ile Asn Ser Leu Gly Leu Thr Asp Gln Asp Leu Asp Phe Ile Asn Phe Asp Leu Thr Asn Lys Lys His Thr Pro Arg Ile Ala Ala   116
831 ATC AAC TCA TTA GGG TTA ACA GAT CAA GAT TTG GAT TTT ATT AAT TTT GAT TTA ACC AAC AAA AAA CAT ACA CCA AGA ATC GCA GCC
    His Tyr Asp His Tyr Ser Asp Val Leu Thr Lys Phe Gly Asp Arg Leu Lys Ser Glu Cys Ala Lys Asp Ser Phe Gly Asn Ala Val   145
918 CAT TAC GAT CAC TAT TCT GAT GTT CTA ACT AAG TTT GGC GAT CGA CTC AAA AGT GAA TGT GCA AAA GAC TCT TTT GGG AAT GCA GTG
    Glu Thr Lys Asn Gly Gln Ile Gln Thr Trp Leu Leu Tyr Asn Asp Lys Ile Tyr Cys Ser Ala Asn Asp Leu Phe Ala Leu Arg Thr   174
1005 GAA ACG AAA AAT GGT CAA ATT CAA ACG TGG TTA CTA TAT AAC GAT AAG ATA TAT TGT TCG GCT AAT GAT TTG TTT GCA TTA CGA ACT
    Asp Leu Ser Ser His Ser Thr Leu Leu Phe Asp Arg Ile Ile Gly Lys Ser Lys Asp Ala Pro Leu Val Ile Leu Tyr Gly Ser Pro   203
1092 GAT TTG AGT TCT CAT TCT ACA CTT TTA TTT GAT AGG ATT ATT GGA AAA TCA AAA GAT GCA CCT TTG GTG ATT TTA TAT GGA AGC CCG
    Thr Glu Glu Leu Thr Lys Asp Phe Leu Lys Ile Leu Tyr Pro Asp Ala Lys Ala Gly Lys Leu Lys Phe Val Trp Arg Tyr Ile Pro   232
1179 ACT GAG GAA CTG ACT AAA GAT TTT CTT AAA ATA TTG TAT CCA GAT GCA AAG GCT GGA AAA TTA AAG TTT GTA TGG AGG TAC ATT CCA
    Leu Gly Ile Lys Lys Leu Asp Ser Ile Ser Gly Tyr Gly Val Ser Leu Lys Met Glu Lys Tyr Asp Tyr Ser Gly Ala Glu Gly Asn   261
1266 CTG GGA ATC AAA AAA CTG GAC TCA ATT TCT GGA TAC GGT GTA TCA TTG AAA ATG GAA AAG TAT GAT TAT TCT GGT GCA GAA GGA AAT
    Pro Lys Tyr Asp Leu Ser Arg Asp Phe Thr Arg Ile Asn Asp Ser Gln Glu Leu Val Leu Val Asn Glu Lys His Ser Tyr Glu Leu   290
1353 CCA AAG TAT GAT TTG AGT CGA GAT TTC ACC AGA ATT AAT GAC TCG CAA GAG TTG GTC CTG GTC AAT GAA AAA CAT TCG TAT GAA CTT
    Gly Val Lys Leu Thr Ser Phe Ile Leu Ser Asn Arg Tyr Lys Ser Thr Lys Tyr Asp Leu Leu Asp Thr Ile Leu Thr Asn Phe Pro   319
1440 GGT GTT AAA TTG ACT TCA TTC ATA TTA TCC AAT CGT TAC AAG AGT ACT AAA TAT GAC CTT TTA GAT ACG ATT TTA ACC AAC TTT CCC
    Lys Phe Ile Pro Tyr Ile Ala Arg Leu Pro Lys Leu Leu Asn His Glu Lys Val Lys Ser Lys Val Leu Gly Asn Glu Asp Ile Gly   348
1527 AAG TTT ATT CCT TAC ATT GCA CGA TTA CCA AAA TTA CTA AAT CAT GAA AAA GTT AAA TCC AAA GTG CTT GGA AAT GAA GAT ATA GGG
    Leu Ser Gln Asp Ser Tyr Gly Ile Tyr Ile Asn Gly Ser Pro Ile Asn Pro Leu Glu Leu Asp Ile Tyr Asn Leu Gly Thr Arg Ile   377
1614 CTA TCT CAA GAC TCC TAC GGA ATA TAT ATC AAC GGT TCC CCA ATA AAT CCA CTA GAG TTA GAT ATT TAC AAT CTA GGT ACC AGG ATA
    Lys Glu Glu Leu Gln Thr Val Lys Asp Leu Val Lys Leu Gly Phe Asp Thr Val Gln Ala Lys Leu Leu Ile Ala Lys Phe Ala Leu   406
1701 AAG GAG GAA TTA CAG ACT GTG AAA GAT TTA GTG AAA CTT GGA TTT GAT ACC GTA CAA GCA AAG CTC TTG ATA GCA AAA TTT GCT TTA
    Leu Ser Ala Val Lys Gln Thr Gln Phe Arg Asn Gly Asn Thr Leu Met Gly Asn Asn Glu Asn Arg Phe Lys Val Tyr Glu Asn Glu   435
1788 CTT TCA GCT GTT AAA CAA ACA CAA TTT CGA AAT GGG AAT ACA TTA ATG GGT AAC AAT GAA AAT AGA TTT AAA GTG TAT GAA AAT GAA
    Phe Lys Lys Gly Ser Ser Glu Lys Gly Gly Val Leu Phe Phe Asn Asn Ile Glu Leu Asp Asn Thr Phe Lys Glu Tyr Thr Thr Asp   464
1875 TTT AAG AAG GGT AGT TCA GAA AAG GGT GGG GTC TTG TTT TTC AAT AAC ATT GAA TTA GAC AAC ACA TTC AAG GAG TAC ACC ACT GAT
    Arg Glu Glu Ala Tyr Leu Gly Val Gly Ser His Lys Leu Lys Pro Asn Gln Ile Pro Leu Leu Lys Glu Asn Ile His Asp Leu Ile   493
1962 CGT GAG GAG GCA TAT TTA GGA GTT GGT TCT CAT AAA CTT AAG CCA AAT CAA ATT CCG TTA TTG AAA GAG AAC ATC CAT GAT TTA ATT
    Phe Ala Leu Asn Phe Gly Asn Lys Asn Gln Leu Arg Val Phe Phe Thr Leu Ser Lys Val Ile Leu Asp Ser Gly Ile Pro Gln Gln   522
2049 TTC GCA TTA AAT TTT GGG AAC AAA AAC CAA TTG CGG GTG TTT TTC ACT TTA TCT AAG GTG ATT TTG GAC TCC GGT ATA CCT CAA CAA
    Val Gly Val Leu Pro Val Ile Gly Asp Asp Pro Met Asp Leu Leu Leu Ala Glu Lys Phe Tyr Trp Ile Ala Glu Lys Ser Ser Thr   551
2136 GTT GGA GTT TTG CCC GTT ATA GGA GAT GAC CCA ATG GAT CTG TTA CTC GCT GAG AAA TTT TAT TGG ATT GCT GAG AAA TCA AGC ACA
    Gln Glu Ala Leu Ala Ile Leu Tyr Lys Tyr Phe Glu Ser Asn Ser Pro Asp Glu Val Asp Asp Leu Leu Asp Lys Val Glu Val Pro   580
2223 CAA GAG GCA TTA GCA ATA TTG TAT AAA TAT TTT GAA TCA AAC AGT CCA GAT GAA GTT GAT GAC TTA TTA GAT AAA GTG GAA GTA CCC
    Glu Asp Tyr Lys Val Asp Tyr Asn His Val Leu Asn Lys Phe Ser Ile Ser Thr Ala Ser Val Ile Phe Asn Gly Val Ile Tyr Asp   609
2310 GAA GAT TAT AAA GTC GAT TAT AAT CAT GTG TTA AAC AAG TTT TCT ATA TCA ACT GCT TCG GTC ATT TTC AAT GGG GTT ATT TAC GAT
    Leu Arg Ala Pro Asn Trp Gln Ile Ala Met Ser Lys Gln Ile Ser Gln Asp Ile Ser Leu Ile Lys Thr Phe Leu Arg Gln Gly Pro   638
2397 TTA AGA GCA CCA AAC TGG CAG ATT GCA ATG AGT AAA CAA ATA TCC CAG GAC ATT TCA CTT ATT AAA ACT TTC TTG AGA CAG GGA CCA
    Ile Glu Gly Arg Leu Lys Asp Val Leu Tyr Ser Asn Ala Lys Ser Glu Arg Asn Leu Arg Ile Ile Pro Leu Glu Pro Ser Asp Ile   667
2484 ATA GAG GGT AGA TTG AAA GAT GTT CTT TAC TCT AAT GCA AAA TCA GAA CGC AAT TTA CGT ATA ATT CCA TTA GAA CCT AGT GAC ATT
    Ile Tyr Lys Lys Ile Asp Lys Glu Leu Ile Asn Asn Ser Ile Ala Phe Lys Lys Leu Asp Lys Ala Gln Gly Val Ser Gly Thr Phe   696
2571 ATT TAC AAG AAA ATC GAC AAG GAA TTA ATA AAC AAT TCA ATT GCA TTC AAG AAG CTA GAT AAA GCG CAG GGT GTG TCT GGA ACA TTT
```

Figure 1 A (continued)

Trp Leu Val Ser Asp Phe Thr Lys Ser Ala Ile Ile Thr Gln Leu Ile Asp Leu Leu Leu Leu Leu Lys Lys Lys Ala Ile Gln Ile 725

2658 TGG CTA GTG TCG GAT TTT ACC AAG TCA GCA ATA ATT ACT CAA TTG ATA GAT TTG TTA TTG CTT CTC AAA AAG AAA GCA ATT CAG ATA

Arg Ile Ile Asn Thr Gly Asp Thr Asp Val Phe Gly Lys Leu Lys Thr Lys Phe Lys Leu Thr Ala Leu Thr Asn Gly Gln Ile Asp 754

2745 AGA ATT ATT AAT ACT GGG GAT ACA GAT GTT TTT GGA AAA TTG AAA ACA AAG TTT AAA TTA ACC GCC TTA ACA AAT GGA CAA ATT GAT

Glu Ile Ile Glu Ile Leu Lys Lys Ser Asn Ala Ser Ser Ala Asn Asn Asp Glu Leu Lys Lys Met Leu Glu Thr Lys Gln Leu Pro 783

2832 GAA ATT ATT GAG ATT TTG AAA AAA TCC AAC GCT TCA AGT GCA AAT AAT GAT GAA TTG AAA AAA ATG CTT GAG ACT AAG CAA TTA CCT

Ala His His Ser Phe Leu Leu Phe Asn Ser Arg Tyr Phe Arg Leu Asp Gly Asn Phe Gly Tyr Glu Glu Leu Asp Gln Ile Ile Glu 812

2919 GCT CAT CAC TCT TTT TTG CTA TTC AAC TCT AGA TAT TTT AGA TTG GAT GGA AAT TTT GGA TAC GAG GAA TTG GAT CAA ATT ATA GAG

Phe Glu Val Ser Gln Arg Leu Asn Leu Ile Pro Asp Ile Met Glu Ala Tyr Pro Asp Glu Phe Arg Ser Lys Lys Val Ser Asp Phe 841

3006 TTT GAA GTA TCT CAA AGA TTG AAC TTA ATC CCG GAC ATC ATG GAG GCA TAT CCG GAT GAG TTT AGG TCG AAG AAG GTA AGT GAT TTT

Asn Leu Val Leu Ser Gly Leu Asp Asn Met Asp Trp Phe Asp Leu Val Thr Ser Ile Val Thr Lys Ser Phe His Val Asp Glu Lys 870

3093 AAT CTG GTT TTG TCT GGA TTA GAC AAT ATG GAC TGG TTT GAT TTG GTG ACT TCC ATA GTG ACA AAA TCA TTC CAT GTC GAC GAA AAA

Arg Phe Ile Val Asp Val Asn Arg Phe Asp Phe Ser Ser Leu Asp Phe Ser Asn Ser Ile Asp Val Thr Thr Tyr Glu Glu Asn Ser 899

3180 AGG TTT ATT GTT GAT GTT AAC AGG TTT GAT TTT AGC TCA TTG GAT TTT TCA AAC TCG ATT GAT GTA ACG ACT TAT GAA GAA AAT AGT

Pro Val Asp Val Leu Ile Ile Leu Asn Pro Met Asp Glu Tyr Ser Gln Lys Leu Ile Ser Leu Val Asn Ser Ile Thr Asp Phe Leu 928

3267 CCA GTT GAT GTA TTA ATA ATT TTG AAC CCT ATG GAT GAA TAT TCT CAA AAA TTG ATA AGC CTT GTT AAT AGC ATT ACA GAT TTT CTG

Phe Leu Asn Ile Arg Ile Leu Leu Gln Pro Arg Val Asp Leu Lys Glu Glu Ile Lys Ile His Lys Phe Tyr Arg Gly Val Tyr Pro 957

3354 TTC TTG AAC ATT AGA ATC TTA CTA CAA CCA AGA GTG GAT CTG AAA GAA GAG ATC AAA ATT CAC AAG TTT TAT CGT GGT GTG TAT CCT

Gln Pro Thr Pro Lys Phe Asp Ser Asn Gly Lys Trp Ile Gln His Tyr Ser Ala Gln Phe Glu Ser Ile Pro Ser Asn Val Thr Tyr 986

3441 CAA CCG ACT CCC AAA TTT GAT TCC AAT GGC AAG TGG ATC CAA CAT TAT TCA GCT CAA TTT GAA AGT ATT CCA TCC AAT GTG ACC TAT

Ser Thr Glu Leu Asp Val Pro His Lys Trp Ile Val Val Pro Gln Leu Ser Ser Met Asp Leu Asn Thr Ile Asn Phe Ser Glu Ser 1015

3528 TCT ACT GAA TTA GAT GTT CCA CAT AAG TGG ATA GTT GTT CCT CAA CTG AGT TCG ATG GAT TTA AAC ACA ATC AAT TTC AGC GAA AGC

His Ser Val Asp Ala Lys Tyr Ser Leu Lys Asn Ile Leu Ile Glu Gly Tyr Ala Arg Asp Ile His Thr Gly Lys Ala Pro Asp Gly 1044

3615 CAC TCT GTT GAT GCA AAA TAC TCT CTA AAA AAT ATA TTA ATT GAA GGA TAT GCT AGA GAT ATT CAT ACT GGG AAG GCC CCT GAT GGT

Leu Ile Phe Arg Ala Phe Asn Lys Asn Tyr Ser Thr Asp Thr Leu Val Met Thr Ser Leu Asp Tyr Phe Gln Ile Lys Ala Tyr Pro 1073

3702 TTA ATC TTT AGA GCC TTT AAT AAA AAT TAC TCA ACT GAT ACT TTG GTG ATG ACT TCC TTG GAC TAT TTT CAA ATC AAA GCG TAT CCT

Ser Ile Phe Asn Phe Ser Thr Thr Ser Asn Asp Thr Leu Leu Ser Ala Ser Glu Asn Lys Tyr Gln Ala Asn Thr Glu Glu Leu Glu 1102

3789 AGT ATT TTC AAC TTT AGT ACG ACC TCA AAT GAC ACA TTA TTG TCT GCA TCG GAA AAC AAA TAT CAG GCT AAT ACC GAG GAA TTG GAG

Ser Ile Glu Val Pro Val Phe Lys Ile Asp Gly Ser Thr Ile Tyr Pro Arg Val Met Lys Ser Gly Asn Asn Lys Pro Met Leu Thr 1131

3876 AGC ATT GAG GTG CCA GTT TTT AAA ATT GAT GGA TCG ACC ATA TAT CCA AGG GTA ATG AAA TCT GGC AAC AAT AAG CCA ATG CTG ACG

Arg Lys His Ala Asp Ile Asn Ile Phe Thr Ile Ala Ser Gly Gln Leu Tyr Glu Lys Leu Thr Ser Ile Met Ile Ala Ser Val Arg 1160

3963 AGA AAA CAT GCA GAT ATA AAT ATT TTT ACA ATT GCT AGT GGC CAA CTT TAT GAA AAG TTA ACT AGC ATT ATG ATT GCG TCA GTA AGA

Lys His Asn Pro Ser Leu Thr Ile Lys Phe Trp Ile Leu Glu Asp Phe Val Thr Pro Gln Phe Lys His Leu Val Glu Leu Ile Ser 1189

4050 AAA CAT AAC CCT AGC CTG ACA ATA AAA TTC TGG ATT TTG GAA GAT TTT GTG ACC CCA CAA TTC AAA CAC TTG GTA GAG CTT ATC TCA

Ile Lys Tyr Asn Val Glu Tyr Glu Phe Ile Ser Tyr Lys Trp Pro Asn Phe Leu Arg Lys Gln Lys Thr Lys Glu Arg Met Ile Trp 1218

4137 ATA AAG TAT AAT GTC GAA TAT GAG TTT ATT AGT TAC AAA TGG CCC AAT TTC TTG AGA AAA CAG AAA ACC AAA GAA AGA ATG ATT TGG

Gly Tyr Lys Ile Leu Phe Leu Asp Val Leu Phe Pro Gln Asp Leu Asn Lys Ile Ile Phe Ile Asp Ala Asp Gln Ile Cys Arg Ala 1247

4224 GGG TAT AAG ATT TTG TTT TTG GAC GTT TTG TTC CCA CAA GAT CTC AAC AAG ATT ATA TTC ATT GAC GCC GAT CAA ATA TGT AGG GCA

Asp Leu Thr Glu Leu Val Asn Met Asp Leu Glu Gly Ala Pro Tyr Gly Phe Thr Pro Met Cys Asp Ser Arg Glu Glu Met Glu Gly 1276

4311 GAT TTG ACA GAA TTG GTT AAC ATG GAT CTT GAA GGT GCT CCA TAT GGA TTT ACT CCA ATG TGT GAT TCT CGG GAA GAA ATG GAA GGT

Phe Arg Phe Trp Lys Glu Gly Tyr Trp Ser Asp Val Leu Lys Asp Asp Leu Lys Tyr His Ile Ser Ala Leu Phe Val Val Asp Leu 1305

4398 TTC AGA TTT TGG AAA GAA GGA TAC TGG TCC GAT GTT TTG AAG GAT GAT TTG AAA TAT CAT ATT AGT GCA TTA TTT GTT GTT GAT TTG

Gln Lys Phe Arg Ser Ile Lys Ala Gly Asp Arg Leu Arg Ala His Tyr Gln Lys Leu Ser Ser Asp Pro Asn Ser Leu Ser Asn Leu 1334

4485 CAA AAG TTC AGA TCT ATA AAA GCT GGA GAC AGA TTG AGA GCA CAC TAT CAA AAG CTT TCT AGT GAT CCA AAT TCG TTG AGC AAT TTA

Asp Gln Asp Leu Pro Asn Asn Met Gln Arg Leu Ile Lys Ile Phe Ser Leu Pro Gln Asn Trp Leu Trp Cys Glu Thr Trp Cys Ser 1363

4572 GAT CAA GAT TTG CCC AAT AAT ATG CAA AGA CTG ATA AAA ATT TTC AGT TTG CCT CAA AAT TGG CTC TGG TGT GAA ACG TGG TGC TCA

Asp Lys Ser Leu Glu Asp Ala Lys Met Ile Asp Leu Cys Asn Asn Pro Leu Thr Arg Glu Asn Lys Leu Asp Ala Ala Lys Arg Leu 1392

4659 GAT AAA AGC TTG GAA GAT GCA AAA ATG ATT GAT CTT TGC AAC AAT CCA TTA ACT AGA GAA AAT AAA TTA GAT GCT GCT AAG AGA TTG

Ile Pro Glu Trp Ile Glu Tyr Glu Gln Glu Ile Glu Pro Leu Val Ser Leu Val Gln Asn Asn Thr Ala Lys Glu Val Val Gln Glu 1421

4746 ATC CCA GAA TGG ATT GAA TAC GAG CAA GAA ATT GAA CCA TTG GTA TCA TTA GTA CAG AAT AAT ACC GCC AAA GAA GTT GTT CAA GAG

Ile Glu Ile Asp Thr Asp Gly Glu Gln Glu Glu Gln Lys Gln Glu Ser Asn Asp Asp Asp Phe Ile **His Asp Glu Leu** stop 1447

4833 ATA GAA ATT GAT ACA GAC GGA GAA CAA GAA GAA CAA AAA CAA GAA AGT AAT GAT GAT GAT TTT ATT CAC GAT GAA TTG TAA TTGTCAA

4921 AGTCACATGGAATAAATAGTGAGAACTCCTGAAACGGCATTAAATACGCACGTTGGGTAGAGATAATACAAATATAGATAAATAGATAGAGAGAAAAAAATGTTGGATTTTTTTC

5036 AGACTTCTCTTCCTCCTGGGCGCCTCCGGTTTAACTATAATTTTTTAAGATTACACAAAATTCAAGTACACGCACTTTCTAATTATTTTATTGAAGAGTCATAATCAGTAATGAA

Figure 1B.

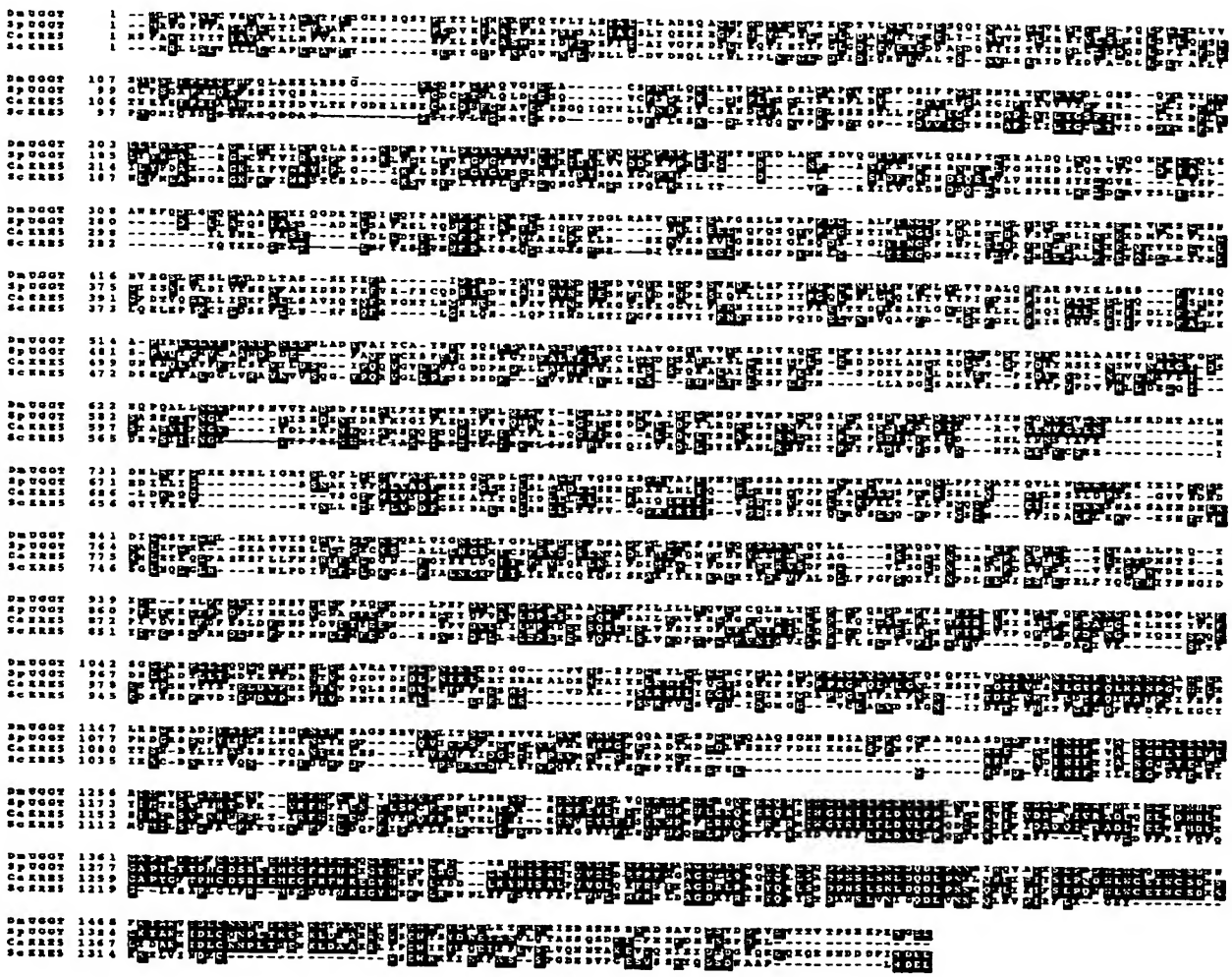


Figure 2A.

TTTTNGCCCCATTCAGTATT

1

21 CAGTATTNTATATAATATATAATATATAAACATAACATATTAAAAAAAACTTCATATTTTCAACTGTTCAACCCAACACCCCCCTGTTCCTAGTCTAGTCCAATTTATTAAGTCA

136 TATATTGTTGATTAATTAAACAGTCACTAGTCCTCAGTCCTCAGTCCTTAGTTCTTGGTTCTTAATATTAAGATNTTCCATTTTTTTTTTTTTACCCAAGCTATGAAAATTATTTT

251 TGTTGTCTAACAACTATAATAATATTTACCAGAAATTGCTACAAATATAAATAAATAAATAAATAAATATAATTAAGAGTATATCTCCCCTTTTGTTTTTTTTTTCTTCCCAGCC

Met Ser Asp Ser Glu Ser Tyr Tyr Gln Asn Ser Thr Thr Asn Gln Pro Ile Pro Arg Ser Asp Glu Val Leu Asp Asp His Arg Asn 29

366 ATG TCC GAT AGT GAA AGT TAT TAT CAA AAT TCA ACT ACT AAT CAA CCT ATT CCT AGA TCT GAT GAA GTA TTG GAT GAT CAT AGA AAT

Gln Ile Thr Asn Asp Cys Ala Ile Ser Asp Ser Glu Asp Glu Leu Glu Leu Lys Ser Glu Leu Glu Ser Glu Val Val Lys Ser Glu 58

453 CAA ATC ACT AAT GAT TGT GCC ATT AGT GAT AGT GAA GAT GAG TTG GAA TTA AAA TCA GAA TTA GAA TCA GAA GTT GTA AAA AGC GAA

Lys Gln Gln Gln His His Gln Glu Ile Thr Ser Asp Asn Ala Lys Pro Leu Thr Arg Lys Ser Gly Ser Ser Ile Lys Lys Lys Ser 87

540 AAA CAA CAA CAA CAT CAT CAA GAG ATT ACA TCA GAT AAT GCT AAA CCA TTG ACT CGT AAA TCT GGT TCT TCA ATT AAG AAA AAA TCT

Asn Leu Thr Asp Lys Asp Arg Ile Thr Asn Pro Met Ser Leu Ser Gly Gly Asp Asp Thr Ile Asn Ser Gly His Lys Asn Arg Asn 116

627 AAT CTT ACC GAT AAA GAT AGA ATT ACC AAC CCT ATG AGT TTA TCT GGT GGT GAT GAT ACT ATT AAC AGC GGT CAC AAA AAT CGT AAT

Tyr Asn Met Ser Ser Leu Arg Lys Asp Phe Tyr Leu Lys Asp Asn Thr Asp Asp Asn Ser Thr Asn Asn His Thr His Leu Ala Ile 145

714 TAT AAC ATG AGT TCA TTA CGT AAA GAT TTT TAT TTA AAA GAT AAT ACT GAC GAC AAT TCT ACT AAT AAT CAT ACT CAT CTT GCA ATT

Pro Ile Pro Ile Pro Ile Pro Thr Pro Ile Ile Thr Asn Ala Asn Lys Ser Arg Arg Lys Ser Gln Leu Glu Asn Leu Pro Pro Leu 174

801 CCA ATT CCA ATT CCA ATT CCA ACC CCA ATT ATT ACT AAT GCT AAT AAA TCA AGA AGA AAA TCT CAA TTG GAA AAT TTA CCT CCA TTA

Ile Lys Lys Lys Thr Ile Gly Arg Asn Asn Ser Asn Asn Phe Glu Asn Asp Leu Val Ser Pro Met Thr Lys Met Lys Thr Asn Asp 203

888 ATT AAA AAG AAA ACA ATT GGT CGT AAT AAT TCT AAT AAT TTT GAA AAT GAT TTA GTT AGT CCC ATG ACA AAA ATG AAA ACT AAT GAT

Ser Glu Asp Ile Thr Asn Thr Ser Thr Thr Ala Asn His Met Lys Leu Gly Ile Gly Ala Thr Thr Leu Gly Val Gly Thr Gly Thr 232

975 AGT GAA GAT ATT ACT AAT ACT AGC ACC ACT GCT AAT CAT ATG AAA CTT GGT ATT GGT GCT ACA ACC CTT GGT GTT GGA ACT GGT ACT

Thr Ala Thr Ala Thr Ala Thr Ala Ala Ala Gly Arg Arg Pro Ser Arg Ser Ser Ile Asp Ser Glu Ala Asp Ser His Ala Ser Arg 261

1062 ACC GCC ACT GCC ACT GCC ACT GCT GCT GCT GGT AGA AGA CCA TCT CGT TCA TCT ATT GAT AGT GAA GCT GAT TCT CAT GCA TCA AGA

Ser Ser Gln Glu Thr Glu Glu Asp Val Cys Phe Pro Met Val Gly Asp His Ile Arg Val Asn Gly Ile Asp Phe Asp Glu Ile Asp 290

1149 TCA TCT CAA GAA ACT GAA GAA GAT GTT TGT TTT CCT ATG GTT GGT GAT CAT ATT AGA GTT AAT GGA ATT GAT TTT GAT GAA ATT GAT

Glu Phe Ile Arg Glu Glu Arg Glu Glu Ala Tyr Leu Gln Lys Gln Met Ile Ala Lys Asn Ile Leu Arg Ile Asp Glu Phe Gln Asn 319

1236 GAA TTT ATT AGA GAA GAA AGA GAA GAA GCT TAT TTA CAA AAA CAA ATG ATT GCT AAA AAT ATT CTG CGT ATT GAT GAA TTT CAA AAT

Leu Ser Lys Asn Asn Thr Thr Ser Gly Ala Ser Arg His Pro Tyr His His His Ser Asn Asn Asn Lys Lys Asn Asn Gly Gly Asp 348

1323 CTT TCC AAA AAT AAT ACT ACT AGT GGT GCA TCT CGT CAT CCA TAT CAT CAT CAC AGT AAT AAT AAT AAA AAA AAT AAT GGT GGT GAT

Gly Gly Gly Ser Ser Met Ala Ala Leu Lys Tyr Thr Pro Lys Asn Ile Leu Lys Lys Thr Leu Ser Arg Phe Glu Phe Thr His Glu 377

1410 GGT GGT GGT TCT AGT ATG GCA GCA TTA AAA TAT ACT CCA AAA AAT ATT TTA AAG AAA ACA TTA TCA AGA TTT GAA TTT ACT CAT GAA

Asn Ser Ser Ser Ser Glu Glu Ile Tyr Glu Leu Lys Thr Lys Gln Gln Pro Pro Tyr Lys Tyr Asp Asp Gln Leu Ser Leu Thr Ser 406

1497 AAT TCT TCA TCT TCA GAA GAA ATT TAT GAA TTG AAG ACT AAA CAA CAA CCA CCT TAC AAA TAT GAT GAT CAA TTA TCA TTA ACT TCA

Ser Thr Ser Ser Thr Ser Gly Ser Gly Ser Gly Gln Val Lys Phe Gly Gly Ala Arg Ile Ser Asp Gly Ile Asn Gly Gly Ser Leu 435

1584 TCT ACA TCT TCT ACT TCT GGA TCT GGA TCT GGG CAG GTG AAA TTT GGT GGA GCA AGA ATT TCT GAT GGG ATT AAT GGA GGT TCA TTA

Pro Asp Arg Phe Ser Leu Phe His Ser Glu Ser Glu Glu Thr Ile His Ala Pro Asp Ile Pro Ser Leu Val Ser Pro Gly Gln Ser 464

1671 CCT GAT AGA TTT TCA CTT TTC CAT TCT GAA TCA GAA GAA ACT ATT CAT GCC CCC GAT ATT CCA TCA TTA GTA TCA CCA GGT CAA TCT

Val Arg Asp Leu Phe Arg Asn Gly Glu Glu Thr Trp Trp Leu Asp Cys Thr Cys Pro Thr Asp Ser Glu Met Lys Met Leu Ala Lys 493

1758 GTT CGA GAT TTA TTT AGA AAT GGT GAA GAA ACT TGG TGG TTA GAT TGT ACT TGT CCT ACT GAT TCG GAA ATG AAA ATG TTG GCC AAA

Ala Phe Gly Ile His Pro Leu Thr Ala Glu Asp Ile Arg Met Gln Glu Thr Arg Glu Lys Val Glu Leu Phe Lys Ser Tyr Tyr Phe 522

1845 GCA TTT GGT ATT CAT CCT TTA ACT GCT GAA GAT ATT CGA ATG CAA GAA ACT CGT GAA AAA GTT GAA TTA TTT AAA AGT TAT TAT TTT

Val Cys Phe His Thr Phe Glu Ala Asp Lys Glu Ser Glu Asp Tyr Leu Glu Pro Ile Asn Val Tyr Ile Val Val Phe His Asp Gly 551

1932 GTT TGT TTC CAT ACT TTT GAA GCT GAT AAA GAA TCT GAA GAT TAT TTA GAA CCG ATA AAT GTT TAT ATT GTT GTT TTC CAT GAT GGT

Ile Leu Thr Phe His Phe Ser Pro Ile Ser His Pro Ala Asn Val Arg Arg Arg Val Arg Gln Leu Arg Asp Tyr Val Asp Val Ser 580

2019 ATA TTA ACG TTC CAT TTT TCA CCA ATT TCT CAT CCA GCA AAT GTT AGA AGA AGA GTT CGT CAA TTG AGA GAT TAT GTC GAT GTT AGT

Ala Asp Trp Leu Cys Tyr Ala Leu Ile Asp Glu Ile Thr Asp Gly Phe Ala Pro Val Ile His Gly Ile Glu Tyr Glu Ala Asp Ala 609

2106 GCT GAT TGG TTA TGT TAT GCC TTA ATC GAT GAA ATT ACC GAT GGT TTT GCC CCC GTG ATT CAT GGA ATT GAA TAT GAA GCT GAT GCC

Ile Glu Asp Ala Val Phe Thr Ala Arg Asp Thr Asp Phe Ser Ser Met Leu Gln Arg Ile Gly Glu Ser Arg Arg Lys Val Met Thr 638

2193 ATT GAA GAT GCC GTT TTC ACT GCT AGA GAT ACT GAT TTT AGT AGT ATG TTA CAA AGA ATT GGT GAA TCA AGA AGA AAA GTC ATG ACT

Leu Met Arg Leu Leu Ser Gly Lys Ala Asp Val Ile Lys Met Phe Ala Lys Arg Cys Gln Glu Glu Ala Asn Ser Ser Ser Gly Tyr 667

2280 TTA ATG AGA TTA TTA TCA GGT AAA GCT GAT GTC ATT AAA ATG TTT GCT AAA AGA TGT CAA GAA GAA GCT AAT TCT TCT TCT GGT TAT

Tyr Gln Arg Gln Tyr Asn Leu Gln Gln Gln Gln Gln Gln Ala Pro Pro Pro Pro Pro Asn Pro Ile Ile Thr Ser Pro Ile Asn Ser 696

2367 TAT CAA CGT CAA TAT AAC TTA CAA CAA CAA CAA CAA CAG GCC CCA CCA CCA CCA CCT AAT CCT ATT ATT ACT TCA CCA ATT AAT TCA

Thr Leu Asn Leu Asn Ser Leu Gly Thr Ser Thr Gly Gly Gly Val Gly Val Gly Gly Ile Asn Phe Gly Pro Asn Pro Thr Gly Asn 725

Figure 2 A (continued)

2454 ACT TTG AAT CTT AAT AGT TTA GGA ACT TCA ACT GGT GGA GGA GTA GGA GTA GGA GGA ATT AAT TTT GGT CCC AAT CCA ACT GGA AAT

Asn Thr Asn Thr Asn Thr Asn Thr Thr Gly Ser Pro Ser Pro Pro Gln Gln Gln Gln Gln His Gly Ile Thr Asn Lys Ser Phe Pro 754

2541 AAT ACT AAT ACT AAT ACT AAT ACT ACT GGT TCA CCT TCA CCA CCT CAA CAA CAA CAA CAA CAT GGT ATC ACT AAC AAA TCT TTC CCC

Ile Pro Asp Ala Arg Pro Arg Ala Asp Ile Ala Leu Tyr Leu Gly Asp Ile Gln Asp His Ile Ile Thr Met Phe Gln Asn Leu Leu 783

2628 ATC CCC GAT GCA CGT CCA AGA GCT GAT ATT GCA TTA TAT TTA GGT GAT ATT CAA GAT CAT ATA ATC ACC ATG TTT CAA AAT TTA TTA

Ala Tyr Glu Lys Ile Phe Ser Arg Ser His Ser Asn Tyr Leu Ala Gln Leu Gln Val Glu Ser Phe Asn Ser Asn Asn Lys Ile Thr 812

2715 GCC TAT GAA AAA ATT TTC AGT CGT TCA CAT TCA AAT TAT TTA GCT CAA TTA CAA GTT GAA TCA TTC AAT TCC AAT AAT AAA ATC ACC

Glu Met Phe Ser Lys Ile Thr Leu Ile Gly Thr Met Leu Val Pro Leu Asn Leu Val Thr Gly Leu Phe Gly Met Asn Val Arg Val 841

2802 GAA ATG TTT TCT AAA ATT ACT TTG ATT GGG ACA ATG TTA GTT CCA TTA AAT TTA GTC ACG GGA CTT TTT GGT ATG AAT GTA AGA GTC

Pro Gly Glu Gly Gly Thr Asn Leu Gly Trp Phe Phe Gly Ile Val Gly Val Leu Ile Phe Ile Ile Ile Gly Ser Phe Ile Phe Ala 870

2889 CCT GGT GAA GGT GGT ACC AAT TTA GGT TGG TTT TTC GGA ATT GTT GGA GTA TTA ATA TTT ATA ATT ATT GGA TCA TTT ATA TTT GCT

Gln Trp Trp Leu Lys Lys Leu Asn Asn Ser Ile Glu Gly Gln Asn Asn Gly Asn Arg Pro Ile Phe Asn His Ser Ser Arg Arg Ser 899

2976 CAA TGG TGG TTG AAA AAA TTG AAT AAT TCA ATT GAA GGA CAA AAT AAT GGT AAT CGA CCA ATT TTT AAT CAT TCA TCA AGA AGA TCA

Ile Arg Ser Leu Gly Leu Lys Lys His Gly Gly Asn Lys Ser Ile Ile Ser Phe Pro Asn Lys Tyr Glu Stop 922

3063 ATT AGA AGT TTA GGT TTA AAA AAA CAT GGT GGT AAT AAA TCA ATT ATT AGT TTC CCC AAT AAA TAT GAA TAA GAATAATCAAAGAAATGCC

3154 ACAGAGTTTGATGGTTTGTTTTTTTTTTTTTTATTGTCATGATGGAGTTGTATATACATATACTTTTTATAGAAGTAACAATAGTAAATGATAATAGTAGTCATCAATCATCATA

3269 TTTATAATTGTATATAATCGTATACTAACTTCTTCTTGATTTAGGGAAAGAGTTATATTATTTACTATAAACATTTATTTTTACGAGTTGTCTTAAATTGCAGAGTCAAATTAAT

3384 AGGATGTAAAAGAAGTTTTTAAAGAAGGAATAAAGAAATATTATAATTCAGANGTTCATACAGAAGGGGGGGAAGGAGAAGGGGATATATATCGCCATTTGTTGGTACTTTTGT

3499 TTTTGAAATAAAATATAAGTTTATCTAAATTATTATCAATTATTATCAATATTGCNAATAAATTAACTAATATGACNAATTTTTAAATATGATCCCATTTTTGTTTGAATTGCGA

3614 AGAAAATAAACTTTGAAAACNAATCAGTAAATTCCGTTATATATATTNTNTCC

Figure 2B.

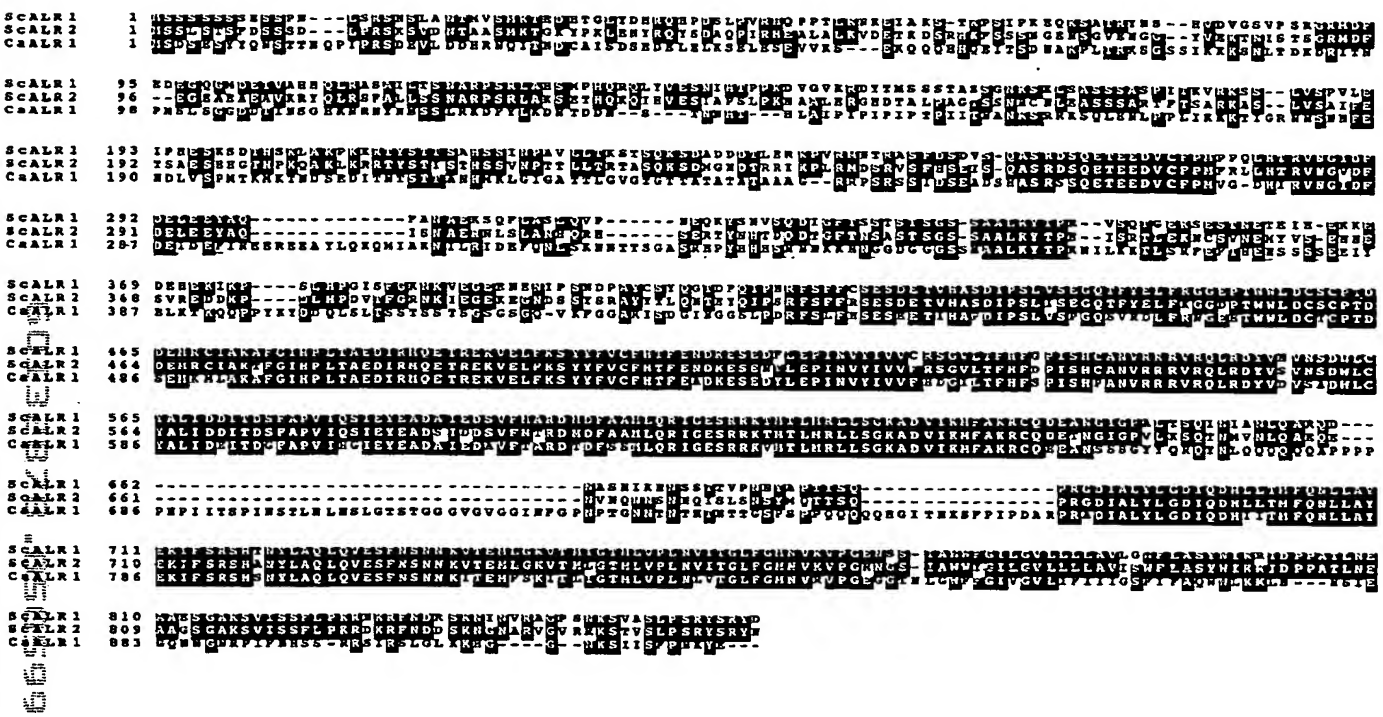


Figure 3A.

1 TTTTCCATTTATTTCCATCTTTCGTTTTCCATTTATTTTATTTTCTTCTTTTTAGTTAGCT

62 CTAATTCAACTTCTTCTACTTCTAACTTGAAATCTAACAACTAAAAAAAGACAAGGAGTGAAAAAAAGTGTGAGAATTTGCTAAAAAAAAATAGAGACAGAAAAAGAAAAAGTTA

177 ACGAACACCAAAGACAGGAAGAAAAAAAAAATTCCAACAACAGGAACAAACATCAACAAACATTAACATCAGCAACAAGCAGAAGACCAATATACATTAACCAATCACACTGAAC

292 TTACTCATAAACTACTTGCTCATATCTTCCTCTTTTTTTTTTTTTTTGCTGCATATTGAAGAAATAGAAACCAAATAGAACCACTCATTATATCTTAATATCAACAAATCCAAAACC

Met Glu His Pro Pro Ala Ala Leu Arg Thr Phe Ser Thr Gln Ser Thr Ser Ser Leu Asn Ser Val Ser Thr Val Ser Ser Ser Arg 29

407 ATG GAA CAT CCA CCA GCA GCT CTC AGA ACA TTT TCA ACC CAA TCA ACT TCA TCT TTG AAT TCA GTA AGT ACT GTT TCG TCT TCA AGA

Ile Val Ser Leu Gly Pro Val Asn Ile Asn Asn Phe Asn Lys Pro Ser Thr Pro Lys Asp His Leu Phe Tyr Arg Cys Glu Ser Leu 58

494 ATT GTT TCT CTG GGC CCA GTC AAT ATA AAC AAT TTC AAT AAA CCA AGT ACT CCC AAA GAC CAT TTA TTC TAT CGA TGT GAA TCA CTA

Lys Arg Lys Leu Gln Lys Ile Pro Gly Met Glu Pro Phe Leu Asn Gln Ala Phe Asn Gln Ala Glu Gln Leu Ser Glu Gln Gln Ala 87

581 AAA CGA AAA CTA CAA AAA ATC CCT GGC ATG GAA CCA TTT TTG AAC CAA GCT TTC AAT CAG GCT GAA CAA CTC AGT GAA CAA CAA GCA

Leu Ala Leu Ala Gln Glu Arg Ser Asn Gly Asn Gly His Ser Asn Gly Lys Arg His Gln Ser Leu Asp Gly Ala Met Asn Arg Leu 116

668 TTG GCT TTG GCA CAG GAA AGA AGC AAT GGA AAT GGA CAT AGT AAT GGC AAA CGT CAT CAA TCA TTA GAC GGT GCC ATG AAT AGA CTT

Ser Val Gly Ser Asp Ser Ser Ser Ile Gln Gly Ser Leu Thr Arg Met Ala Thr Asn Ala Ser Thr Ser Ser Leu Ile Ser Gly Met 145

755 TCA GTT GGT TCT GAT AGT AGT TCG ATT CAA GGT TCA TTG ACA CGA ATG GCT ACC AAT GCG TCA ACG TCA TCT TTA ATC AGT GGT ATG

Pro Asn Ser Asn Thr Leu Phe Thr Phe Thr Ala Gly Val Leu Pro Ala Asn Ile Ser Val Asp Pro Ala Thr His Leu Trp Lys Leu 174

842 CCA AAC AGC AAC ACT TTA TTT ACG TTT ACT GCA GGG GTT TTA CCA GCT AAT ATT AGT GTC GAT CCT GCT ACC CAT CTT TGG AAA TTG

Phe Gln Gln Gly Ala Pro Phe Cys Val Leu Ile Asn His Ile Leu Pro Asp Ser Gln Ile Pro Val Val Ser Ser Asp Asp Leu Arg 203

929 TTC CAA CAA GGG GCC CCC TTT TGT GTT CTT ATC AAT CAT ATC CTT CCT GAT TCC CAA ATA CCA GTT GTC AGT TCT GAT GAC TTG AGA

Ile Cys Lys Lys Ser Val Tyr Asp Phe Leu Ile Ala Val Lys Thr Gln Leu Asn Phe Asp Asp Glu Asn Met Phe Thr Ile Ser Asn 232

1016 ATT TGC AAA AAA TCA GTA TAT GAC TTT TTA ATT GCC GTC AAG ACA CAA TTG AAT TTT GAT GAT GAG AAT ATG TTC ACT ATA TCC AAT

Val Phe Ser Asp Asn Ala Gln Asp Leu Ile Lys Ile Ile Asp Val Ile Asn Lys Leu Leu Ala Glu Tyr Ser Asp Ala Ser Asp Leu 261

1103 GTT TTC TCC GAC AAT GCC CAA GAT TTA ATC AAG ATT ATT GAT GTC ATT AAT AAA CTA CTT GCT GAG TAC TCA GAT GCT AGT GAC CTG

Gly Gly Gly Asp Glu Asp Val Asn Met Asp Val Gln Ile Thr Asp Glu Arg Ser Lys Val Phe Arg Glu Ile Ile Glu Thr Glu Arg 290

1190 GGT GGT GGC GAT GAA GAT GTA AAT ATG GAT GTT CAA ATT ACC GAT GAA AGA TCA AAA GTT TTC CGA GAA ATT ATC GAA ACA GAA AGA

Lys Tyr Val Gln Asp Leu Glu Leu Met Cys Lys Tyr Arg Gln Asp Leu Ile Glu Ala Glu Asn Leu Ser Ser Glu Gln Ile His Leu 319

1277 AAA TAT GTT CAA GAC TTG GAA CTA ATG TGT AAA TAC CGT CAA GAT CTA ATT GAA GCC GAA AAT TTG TCT TCA GAA CAA ATT CAC TTG

Leu Phe Pro Asn Leu Asn Glu Ile Ile Asp Phe Gln Arg Arg Phe Leu Asn Gly Leu Glu Cys Asn Ile Asn Val Pro Ile Arg Tyr 348

1364 TTA TTC CCA AAT TTA AAT GAG ATT ATT GAT TTT CAA AGA CGA TTC CTC AAT GGG TTA GAA TGT AAC ATC AAT GTA CCT ATT AGA TAT

Gln Arg Ile Gly Ser Val Phe Ile His Ala Ser Leu Gly Pro Phe Asn Ala Tyr Glu Pro Trp Thr Ile Gly Gln Leu Thr Ala Ile 377

1451 CAA AGA ATT GGA TCA GTA TTT ATT CAT GCT TCT TTG GGC CCT TTC AAT GCT TAT GAA CCT TGG ACT ATA GGA CAA TTG ACG GCG ATT

Asp Leu Ile Asn Lys Glu Ala Ala Asn Leu Lys Lys Ser Ser Ser Leu Leu Asp Pro Gly Phe Glu Leu Gln Ser Tyr Ile Leu Lys 406

1538 GAT TTG ATC AAC AAA GAA GCT GCT AAT TTG AAA AAA TCG TCA AGT CTA CTT GAT CCT GGG TTT GAA CTT CAA TCG TAT ATA TTA AAG

Pro Ile Gln Arg Leu Cys Lys Tyr Pro Leu Leu Leu Lys Glu Leu Ile Lys Thr Ser Pro Glu Tyr Ser Lys Gln Asp Pro His Gly 435

1625 CCG ATC CAA AGA TTG TGT AAA TAC CCA CTT TTG TTG AAA GAG TTA ATC AAA ACA TCA CCA GAA TAT TCA AAA CAG GAC CCC CAT GGC

Ser Ser Ser Ser Thr Ser Phe Asn Glu Leu Leu Val Ala Lys Thr Ala Met Lys Glu Leu Ala Asn Gln Val Asn Glu Ala Gln Arg 464

1712 AGC TCG TCA TCG ACA TCA TTC AAT GAA TTA TTG GTG GCT AAA ACT GCA ATG AAA GAA TTG GCA AAT CAA GTC AAT GAG GCG CAA AGA

Arg Ala Glu Asn Ile Glu His Leu Glu Lys Leu Lys Glu Arg Val Gly Asn Trp Arg Gly Phe Asn Leu Asp Ala Gln Gly Glu Leu 493

1799 CGA GCA GAA AAT ATC GAA CAT TTG GAA AAA CTA AAA GAA AGA GTA GGT AAT TGG CGT GGG TTT AAT TTG GAT GCT CAA GGA GAA CTA

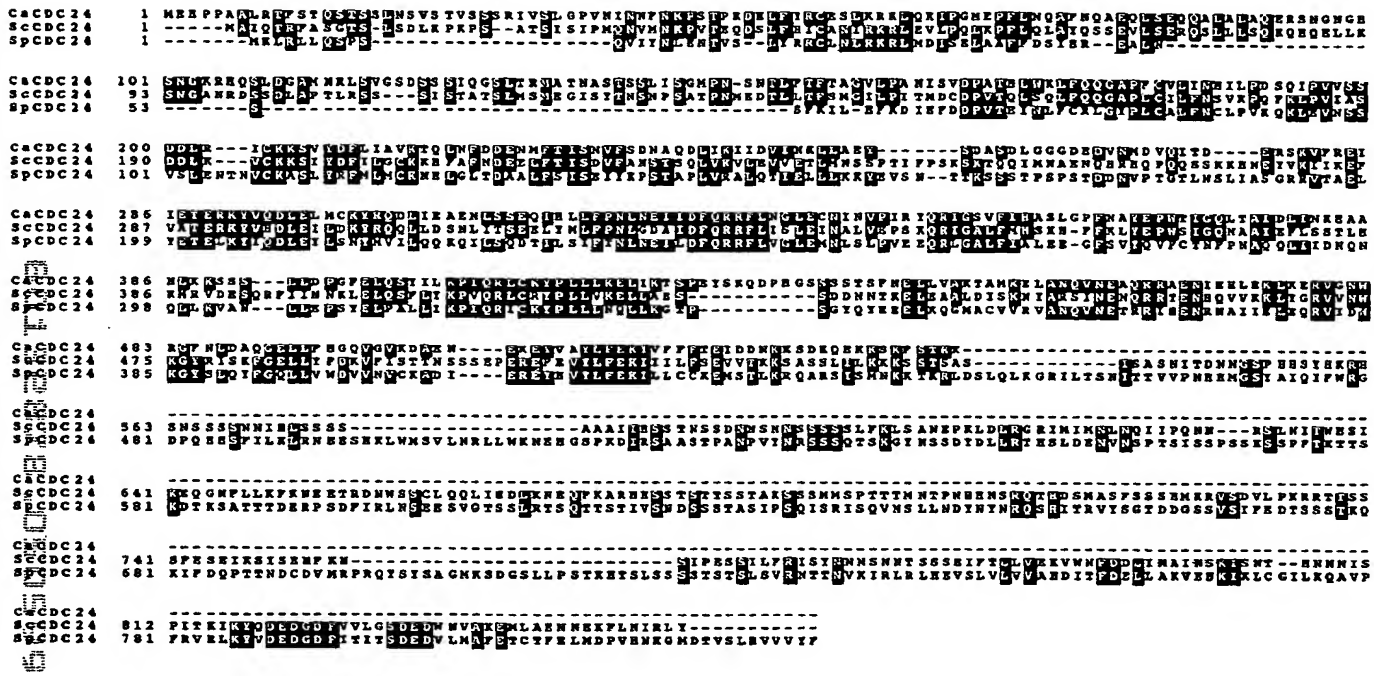
Leu Phe His Gly Gln Val Gly Val Lys Asp Ala Glu Asn Glu Lys Glu Tyr Val Ala Tyr Leu Phe Glu Lys Ile Val Phe Phe Phe 522

1886 TTA TTC CAC GGA CAA GTT GGG GTT AAA GAT GCT GAA AAT GAA AAG GAA TAC GTT GCT TAT CTT TTT GAA AAA ATC GTA TTT TTT TTC

Thr Glu Ile Asp Asp Asn Lys Lys Ser Asp Lys Gln Glu Lys Lys Ser Lys Phe Ser Thr Arg 543

1973 ACA GAA ATT GAT GAT AAC AAA AAA TCT GAT AAA CAG GAA AAG AAG AGC AAG TTT TCG ACA AGA AAG

Figure 3B.



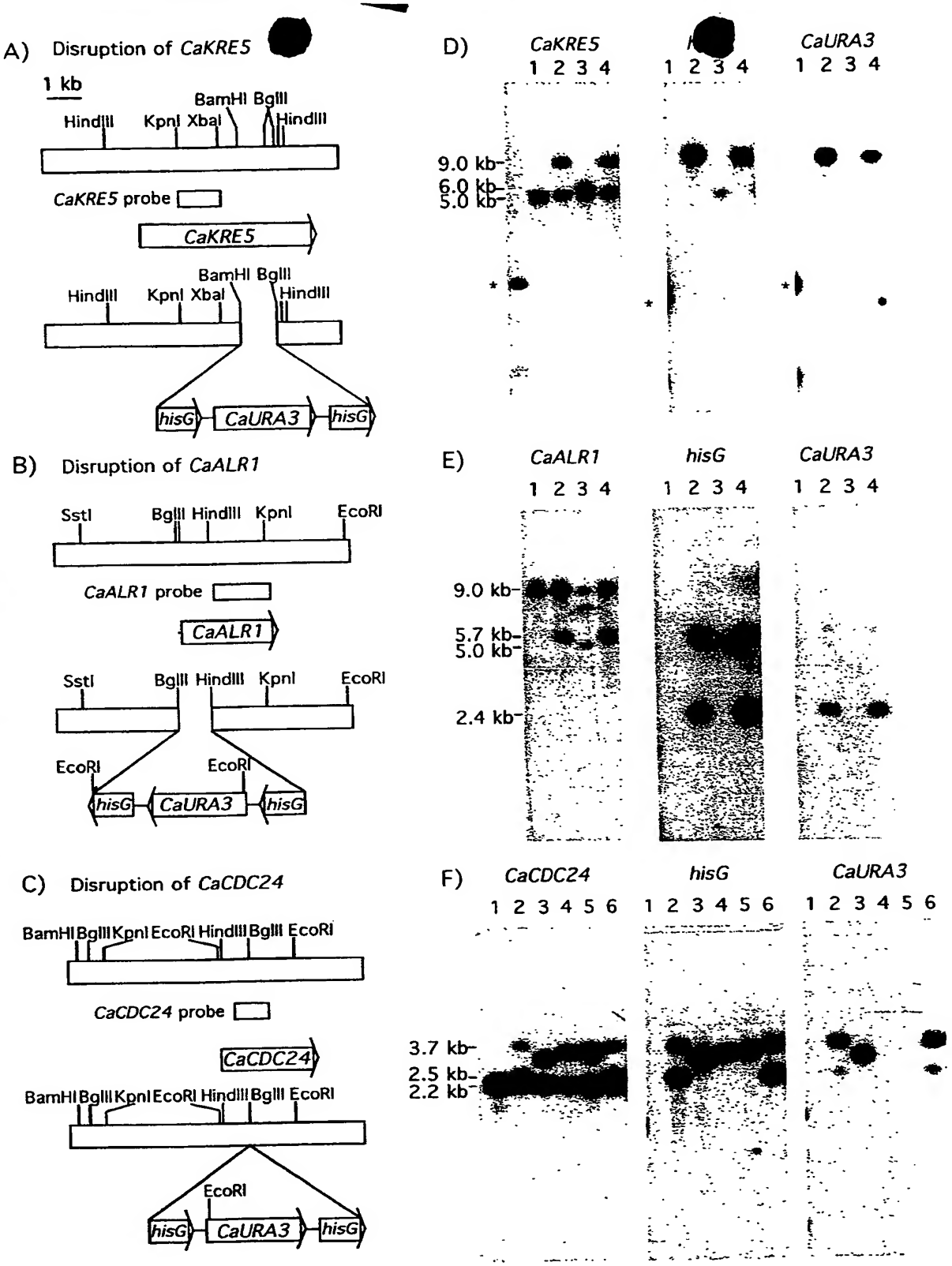

A) Disruption of CaKRE5
1 kb
BamHI BglII
HindIII KpnI XbaI HindIII
CaKRE5 probe
CaKRE5
BamHI BglII
HindIII KpnI XbaI HindIII
hisG CaURA3 hisG
B) Disruption of CaALR1
SstI BglII HindIII KpnI EcoRI
CaALR1 probe
CaALR1
SstI BglII HindIII KpnI EcoRI
EcoRI EcoRI
hisG CaURA3 hisG
C) Disruption of CaCDC24
BamHI BglII KpnI EcoRI HindIII BglII EcoRI
CaCDC24 probe
CaCDC24
BamHI BglII KpnI EcoRI HindIII BglII EcoRI
EcoRI
hisG CaURA3 hisG
D)
CaKRE5
1 2 3 4
1 2 3 4
CaURA3
1 2 3 4
9.0 kb
6.0 kb
5.0 kb
E)
CaALR1
hisG
CaURA3
1 2 3 4
1 2 3 4
1 2 3 4
9.0 kb
5.7 kb
5.0 kb
2.4 kb
F)
CaCDC24
hisG
CaURA3
1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6
3.7 kb
2.5 kb
2.2 kb

Figure 4